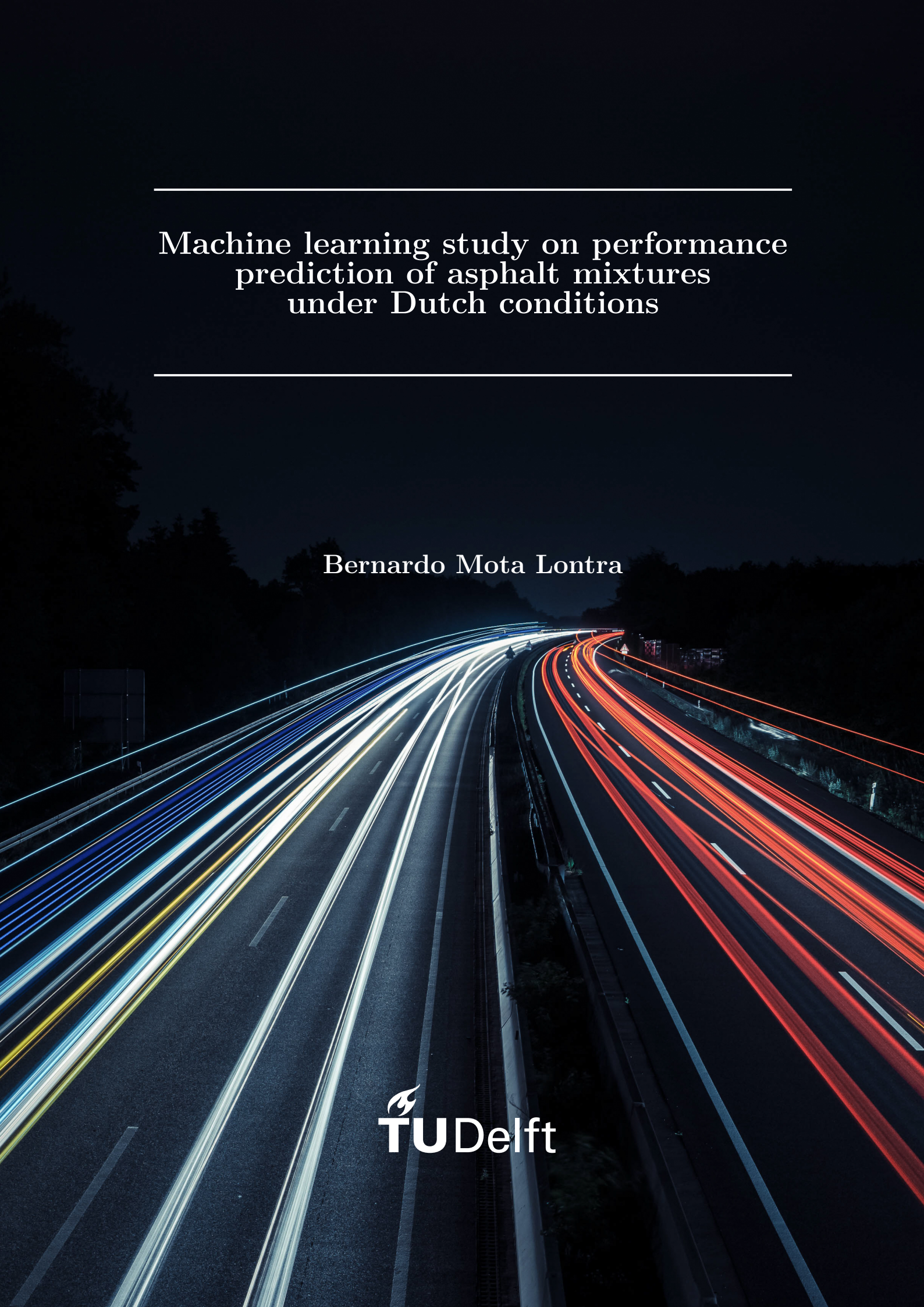
Machine learning study on performance prediction of asphalt mixtures under Dutch conditions

Bernardo Mota Lontra



TU Delft

Machine learning study on performance prediction of asphalt mixtures under Dutch conditions

Master Thesis

Civil Engineering

Structural Engineering track

Student

Bernardo Mota Lontra

5212596

Thesis Committee:	Dr. Kumar Anupam (Chair)
	Prof. dr. ir. Sandra Erkens
	Dr. Alfredo Nuñez
Company Supervisor:	Dr. ir. Mahesh Moenielal

Acknowledgements

The delivery of this report marks the end of my journey through the Technical University of Delft. I want to acknowledge the people involved in the process. I want to thank:

My committee from the TU side. Sandra Erkens, for giving me the opportunity to join this project and valuing my work. The process of working on my thesis and organizing the workshops was a great challenge that helped me improve essential skills that will be needed for future challenges. Kumar Anupam, for receiving me to your team, guiding me and giving me the independence to work. Alfredo Núñez, for accepting the challenge with short notice and being open and flexible in our discussions. This team was an essential pillar for the completion of this project.

The TNO team. Mahesh Moenielal, for the support throughout the project, patience, and guidance. Dave van Vliet for the insightful discussions on the topic, and Almar Snippe for extracting the data required for the work, developing the statistical model, and sharing your expertise. The help given from your side was essential to achieve the end goal.

I cannot forget to thank all the friends I left and made along this journey. Friends who, in the last few years, became my family and welcomed me. Special thanks to Blas, Femke, Francesco, and Matteo, who have been by my side since my first day here and are still by my side. I would also like to thank Chen, Mohammad, and Saranga. All the conversations, coffee breaks, and feedback sessions helped me a lot and motivated me to keep working hard. Thank you for the support and willingness to help organize the workshops.

Por último gostaria de agradecer a minha família. Primeiramente aos meus irmãos que sempre me apoiaram independente da situação e me fizeram acreditar nas minhas escolhas e por isso devo muito a eles. Meus agradecimentos finais vão para Francinete, Gustavo e Mariana. Pessoas que moldaram quem sou e devo literalmente tudo a eles. Obrigado por sempre me proporcionarem mais do que poderia sonhar e por sempre estarem presentes, me guiando e iluminando meu caminho. Muitas vezes duvidei da minha capacidade e pensei em desistir, mas sempre acreditaram e me colocaram de volta no meu lugar. Obrigado novamente por tudo e amo vocês.

Bernardo Mota Lontra

Delft, December 2022

"Quem não ouve a melodia acha maluco quem dança."

Oswaldo Montenegro

Executive summary

The Netherlands is a northwestern European country. The location makes the Netherlands an important entry for trade to the European continent resulting in high traffic intensities [1]. The Netherlands has one of the densest road networks in the world, having over 139.000km of roadways [2, 3].

The understanding of the link between the functional properties and the behavior of the road is crucial [4]. However, predicting such properties is still challenging, and expectations differ from the results found in the field [5, 4].

Data-driven approaches have been part of pavement engineering for decades and have shown to be a powerful tool for performance prediction [6]. This thesis aims to develop a machine learning framework for predicting stiffness, fatigue resistance, resistance to permanent deformation, and water sensitivity. Along with sensitivity analysis to understand the deviations from expectation.

The objective is to develop a framework to understand the applicability of machine learning tools and the impacts of the composition parameters into the mix. The models developed applied three different machine learning tools for regression: Support Vector Machine, Random Forests, and Gradient Boosting. The models were compared to a statistical model to validate the work. The statistical approach was a Multiple Linear Regression. All developed models used the database from the NL-LAB project.

The machine learning models were compared, and the best-performing proceeded to a sensitivity analysis. The sensitivity analysis used SHAP values, which derive from the Games Theory and have shown to be powerful tools for complex model interpretability [7].

The models had good accuracy prediction. For most properties, machine learning had a higher performance than the statistical model. Gradient Boosting performed the best from the machine learning tools and was selected to finalize the research. The sensitivity analysis had good results, confirming some of the expectations and setting a precedent for researching others.

Contents

List of Figures

List of Tables

1

Introduction

This chapter aims to introduce the topic discussed in this thesis, the motive, formulate the research questions and describe the methodology applied to answer them.

1.1 Motivation

Infrastructure plays an important role in economic growth [8]. It facilitates trades, powers businesses, connects workers to jobs, and creates opportunities for struggling communities [9]. These factors are easily visible in the Netherlands, which has over 139.000km of roadways and is heavily dependent on its transport and logistics [10, 11].

The road network in the Netherlands is assessed through the functional properties of asphalt concrete [4]. Although the link between the functional properties and behavior of the road is crucial, the trends expected in properties appear to deviate from expectations [4]. Moreover, prediction of these properties is still challenging due to the complexity of its behavior for cyclic effects and systematic changes throughout its life cycle [5].

Data analyses have been part of modern pavement engineering since the early stages [6]. Different machine learning models were applied to identify critical features and evaluate pavement performance and proved to be powerful tools for predicting [6]. This thesis applies a data-driven approach to developing a framework for performance prediction of asphalt mixtures to address the challenges of functional properties prediction and deviation of expected behavior in the field, focusing on stiffness, fatigue resistance, resistance to permanent deformation, and indirect tensile strength (ITS).

1.2 Problem statement

The scope of this research is to develop a machine learning framework for the prediction of functional properties of asphalt mixtures. The model should incorporate a sensitivity analysis to verify the impact of crucial mixing components. The key components were defined from the NL-LAB project, explained in Chapter 2. The components were the following:

- Bitumen properties
- Bitumen percentage
- Density
- Aging
- Differences between field and laboratory

The developed models focus on stiffness, fatigue resistance, resistance to permanent deformation, and indirect tensile strength prediction using the data from the NL-LAB project. The database includes numerical and descriptive information about the samples. The data does not consider the collected samples' random effects, spacial or traffic information.

Developing a framework requires verifying the applicability of the different tools to the data and its uncertainties. Machine learning models have sets of hyperparameters that, calibrated, improve performance. Defining the optimal hyperparameters and loss function is essential for the framework. Last, it is necessary to validate the final model. Statistical models are well structured and frequently used for several studies and are great candidates for this comparison.

1.3 Research Questions

The available machine learning tools and their variants are vast. Literature review shows that support vector regression, random forests, and gradient boosting are promising tools for this research. Based on the development of the previously described framework, the main research question was formulated as the following:

Does the machine learning framework/models developed in this research improve the performance prediction of asphalt mixtures?

Optimization and validation of the models are necessary to answer the main research question. Multiple linear regression was the selected tool for validating the machine learning models. Therefore, the main question was divided into the following sub-questions:

- How does the machine learning model compare to the statistical model?
- What hyperparameters had the most relevance for the analysis?
- What are the optimal hyperparameter after applying different loss functions?
- Does the machine learning model capture the physical behavior of the mixture?

1.4 Research Methodology

The approach adopted to answer the research questions and understand the effects of crucial mixing components on the functional properties had three main stages. The first part includes the data pre-processing, identification of relevant features, and splitting of the data. Second, the development of the models using different approaches and predictive tools to compare their performance and deal with the non-linearity of the data. In this stage, the models were optimized and cross-validated. Later, a model was selected to proceed to the optimization. After calibrating the model, the third stage was a comparison with the statistical model and sensitivity analysis to identify the impact of the key features.

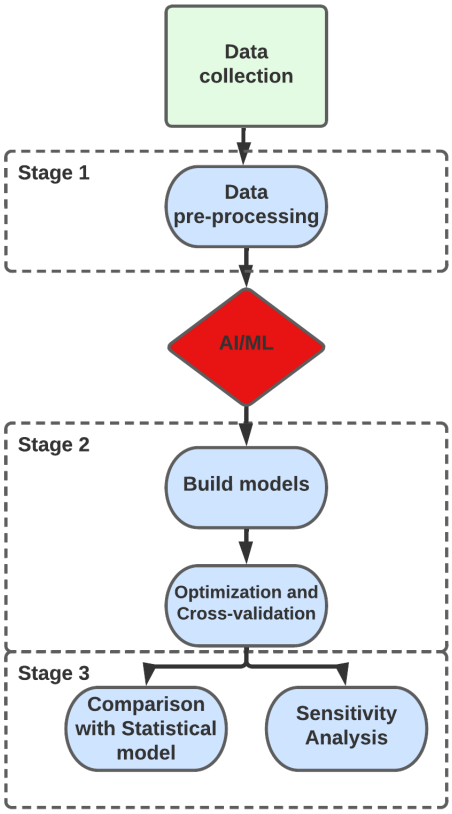


Figure 1.1: Research methodology flowchart

1.5 Thesis Structure

This thesis is divided into five chapters, as illustrated in Figure 1.2. Chapter 1 introduces the topic discussed, motivation, problem statement, formulation of research questions, research methodology, and the structure of the work. Chapter 2 is the literature review, providing background knowledge on the analyzed properties and the models. Chapter 3 introduces the steps and tools used for developing the models and their functionalities. Chapter 4 evaluate the performance of the developed models and perform a sensitivity analysis. Chapter 5 are the conclusions, answering the research questions and giving the recommendations.

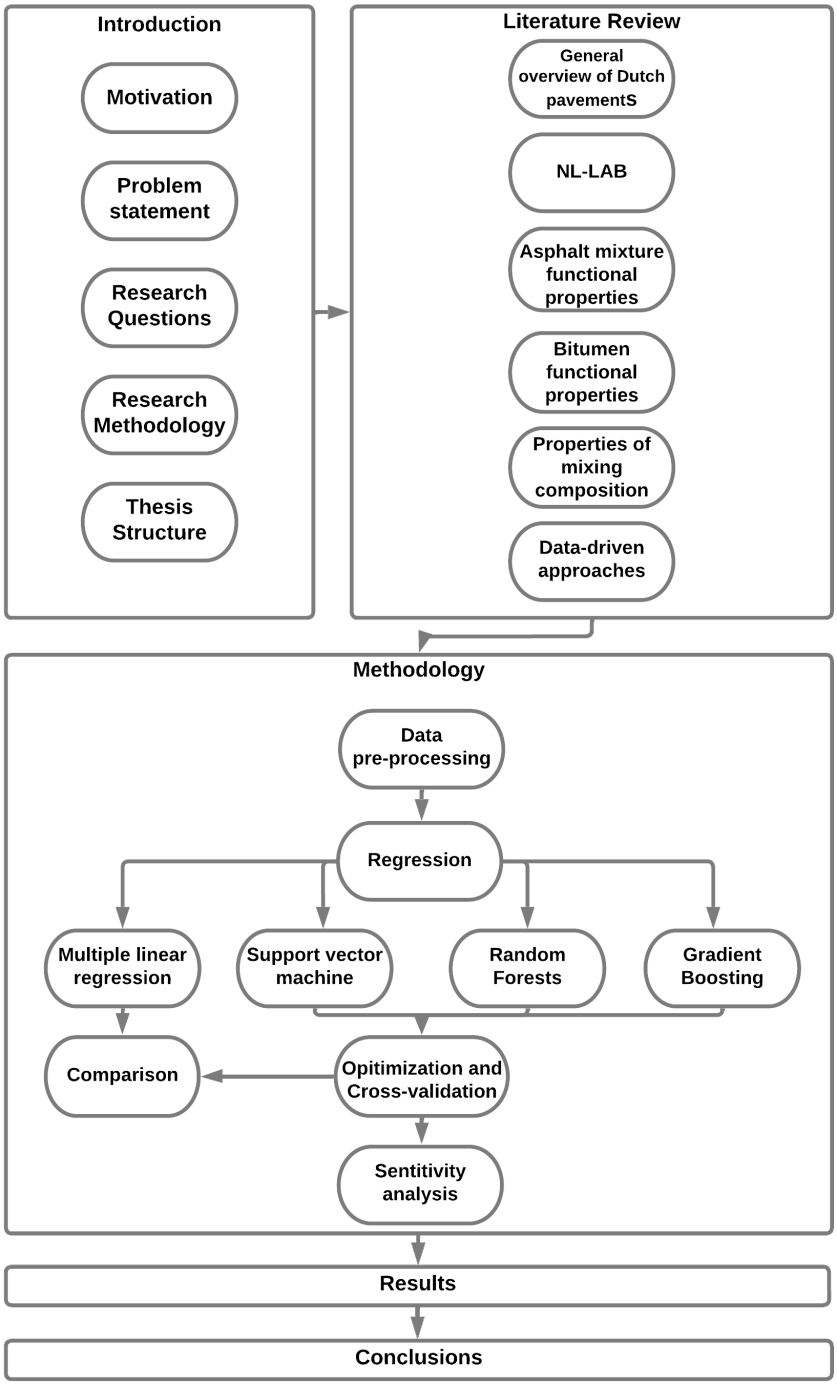


Figure 1.2: Thesis structure flowchart

2

Literature Review

This chapter aims to provide background information on existing literature. (i): An overview of the Dutch pavements and the link to the NL-LAB project. (ii): The background information on the functional properties of asphalt mixtures and binder. (iii): The impacts of the properties of mixing composition in the asphalt mixture. (iv): An overview of the application of data-driven methods in the field.

2.1 General Overview of Dutch pavements

The Dutch road network mainly comprises flexible pavement [3]. Different types of mixtures, such as dense asphalt concrete (DAC), stone mastic asphalt (SMA), and porous asphalt (PA), are used for different specifications [3]. Porous asphalt with two layers (2L-PA) represents more than 80% of the Dutch roads [3]. High precipitation makes roads expected to be wet 13% of the time, which led to the first application of porous asphalt (PA), in 1972 [12]. These mixtures are expected to follow various standards and specifications as specified in CROW(2010) [13].

According to researchers, one of the Netherlands' current challenges is finding the fundamental indicators of in-field pavement performance [1]. It is necessary to cope with the fast changes in mixing compositions caused by the scarcity of raw materials, leading to an increasing necessity for reclaimed, recycled, and bio-based materials. In such case, empirical solutions relations might become obsolete, and time-consuming [1]. As explained in Chapter 1, data-driven approaches have been used in pavement engineering and could be a possible alternative to this challenge.

The scope regarding the challenges in the Netherlands is broad [1]. Companies, research institutes, and the Dutch highway authority, Rijkswaterstaat (RWS), are working on projects to provide solutions. The NL-LAB is one of the projects aiming to assess the functional requirements for pavement performance [1]. The background information regarding NL-LAB is given in the following section.

2.2 NL-LAB

The NL-LAB program uses the Dutch road system as a living laboratory to test materials and monitor their performance [4]. It stands for National or Netherlands Living Lab, and LAB stands for Langjarige Asfalt Bemonstering, which means long-term asphalt sampling in Dutch.

The project started in 2012 when it was realized that some trends expected from experience did not match the ones observed in practice [4]. The initial goal was to verify if the applied functional requirements were a good indicator of the behavior in the field. This data collection also allowed the development of a reference framework for identifying changes and assessing innovations [14].

The data was collected, initially from four different projects, and later,data from more projectes were added. These different projects were named as *Works* in the study [14]. The samples extracted from each Work were from binder/base layers of dense asphalt concrete with Reclaimed Asphalt (RA), and different mixing designs [15]. After two and six years, tests were conducted on the samples and the bitumen at the time of collection. For bitumen, tests were also conducted after six and twelve months to assess the aging [14].

The samples were classified into three different groups. Phase 1, they were mixed and compacted in the lab. In phase 2, the samples were mixed in the field and compacted in the lab. In phase 3, mixing and compaction were done in the field [14]. The samples tested after a few years were kept in the laboratory at the Technical University of Delft at a temperature of 15°C until testing. The data collected over time does not consider the effect of traffic since the focus was only on material properties [4].

The aim was to link the functional properties to materials properties categorized in three main hypotheses, present in Table 2.1 [14]. The first hypothesis is based on the bitumen properties and percentage in the mix. The second hypothesis is related to the density of the samples. The third hypothesis is on the aging of the mixture.

Table 2.1: Main Hypotheses

Hypothesis		Effect	Type test property
1	More and softer bitumen	Lower/Smaller	Stiffness Modulus
		Higher/Larger	Resistance to fatigue
		Lower/Smaller	Resistance to permanent deformation
		Higher/Larger	Water sensitivity
2	Higher density	Higher/Larger	Stiffness Modulus
		Higher/Larger	Resistance to fatigue
		Higher/Larger	Resistance to permanent deformation
		Higher/Larger	Water sensitivity
3	Aging	Higher/Larger	Stiffness Modulus
		Lower/Smaller	Resistance to fatigue
		Higher/Larger	Resistance to permanent deformation
		Lower/Smaller	Water sensitivity

Following tests were carried out for the asphalt concrete mixtures according to the Dutch Standard (RAW 2015):

- Stiffness - EN 12697-26, method B, with four-point bending test on prismatic beams
- Resistance to fatigue - EN 12697-24, method D, with four-point bending test on prismatic beams
- Resistance to permanent deformation - EN 12697-25, method B, cyclical triaxial test with signal
- Water sensitivity - EN 12697-12, method A and EN 12697-23.

Following laboratory tests on bitumen were conducted:

- Bitumen penetration - NEN-EN 1426 - penetration test
- Softening temperature - NEN-EN 1427 - ring and ball test
- Complex shear modulus and phase angle - NEN-EN 14770 (from -10 to 60°C)

Background information on the tests and predictive models to determine the functional properties is presented in the next section.

2.3 Functional properties of asphalt mixture

The determination of the functional properties of asphalt mixture is important in assessing the performance of the overall pavement structure. Different test protocols are used to define this properties. The following section focus on explaining existing predictive models and the norms described in the previous section.

2.3.1 Stiffness

The modulus of elastic materials can be defined by the stress-strain relationship [16]. Visco-elastic materials are frequency and temperature-dependent. Therefore, it is often characterized as a

complex modulus [17]. It can be defined as the relationship between stress and strain for a visco-elastic material [16]. The complex modulus can be divided into a real part that reflects the elastic behavior and an imaginary part that is the internal damping of the material. The phase angle (δ) relates to the damping properties of the material [17]. The stiffness of bituminous mixtures is defined as the absolute value of the complex modulus [17].

$$E^* = |E^*|(cos\delta + isin\delta) \tag{2.1}$$

Predictive models and laboratory tests can predict the stiffness modulus of asphalt mixtures reasonably well [18]. Shell nomograph uses the stiffness of the binder, volume percentages of mineral aggregate, and bitumen to predict stiffness [19]. Witczak model uses a sigmoidal function to fit the material and mixing properties to the stiffness [20]. Micro-mechanical models consider the mechanical and geometrical properties of the constituents to predict stiffness [21]. However, most of them were designed for elastic materials, and several researchers, with limited success, tried translation for visco-elastic materials.

Different tests have been standardized across Europe to define the complex modulus [18]. In the Netherlands, bending tests, direct tensile tests, and indirect tensile tests are used to determine the stiffness modulus [16]. Four-point bending test is one of the commonly used methods to define the stiffness for a required temperature and frequency in prismatic beams [16]. The modulus in the temperature of 20°C and 8 Hz is commonly used in four-point bending [22]. A master curve can be defined to determine the stiffness modulus at an arbitrary temperature, and frequency [16]. The master curve can be fit in a sigmoidal shape for tests performed in at least four temperatures separated by not more than 10°C [16]. The results from tests following the NEN-EN 12697-26 - method B guidelines at 20°C and 8 Hz were used in this research as indicators of stiffness.

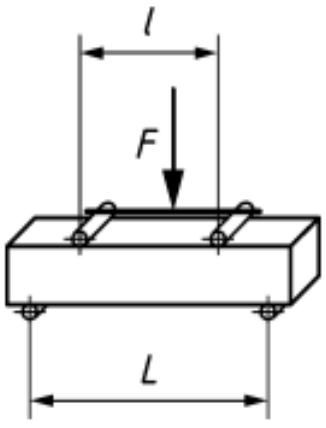


Figure 2.1: Four-point bending test concept [16]

2.3.2 Density

Density is considered one of the most important parameters for pavement construction [23]. The mix needs enough air voids to prevent rutting, water, and air from penetrating through the structure [23]. High air voids lead to water damage, oxidation, and cracking. Lower voids lead to rutting and shoving [23].

In general, when the density of a mixture is increased, the stiffness also increases [4]. However, test results show that a threshold value could be achieved [24]. Tests performed with fine-grade samples, 9.5mm mixtures, showed that, for density levels of 88 and 91%, the change of the stiffness modulus was significant, while 94 and 97% were insignificant. For 12.5mm mixtures, there was still a significant increase at 97% density [24].

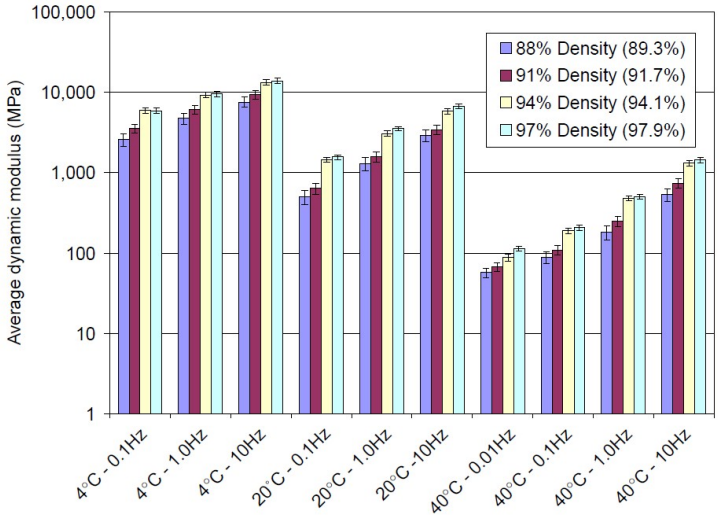


Figure 2.2: Comparison of 9.5 mm dynamic modulus data (|E*|) [24]

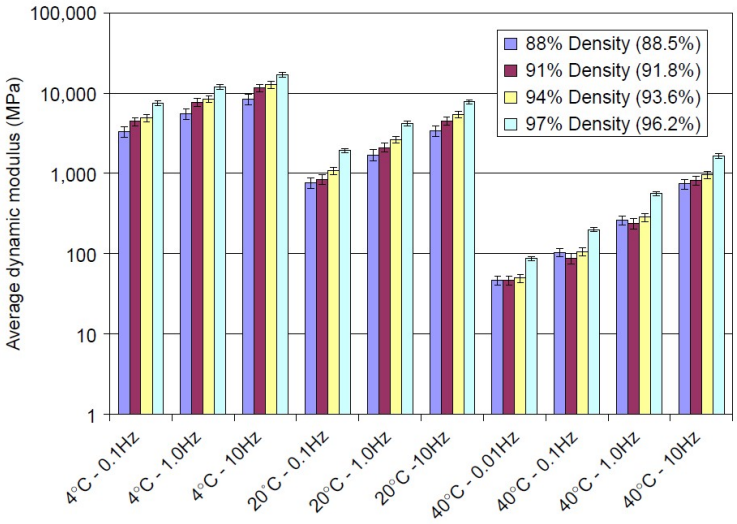


Figure 2.3: Comparison of 12.5 mm dynamic modulus data (|E*|) [24]

Tests to verify the impact of density on fatigue resistance were also performed in 9.5mm and 12.5mm mixtures [24]. The fatigue beam had inconclusive results [24]. Although, samples with 4% air void had longer fatigue life when compared to samples with 6 to 10%. In addition, higher density reduces the air void percentage, which increases the initial stiffness, suggesting a higher fatigue life [25]. The overlay tests showed that improving compaction, and increasing the density, could improve the fatigue performance [24].

A study conducted by Brown and Cross (1989) [26], showed that when the air voids percentage reached 3% or below, rutting was likely to occur. In many cases, insufficient compaction during mix design and testing results in a larger amount of binder required to achieve the expected density [26]. In addition, results from Asphalt Pavement Analyzer (APA) and flow number indicate that higher densities reduce the susceptibility of rutting [24].

The density and air voids are directly correlated to ITS [27]. The increase in void content reduces the ITS because it reduces the cohesion in the mixture. The cohesion is the mechanism that supports the structure for this type of distress.

Performance indicators are also used to assess pavement deterioration over its lifetime. The following section describes the indicators for asphalt mixture performance and the methods for their determination.

2.4 Performance indicators of asphalt mixture

Many factors can cause the deterioration of roads, and indicators of this deterioration, such as resistance to fatigue, resistance to permanent deformation, and water sensitivity, are used to assess the overall performance of the structure.

2.4.1 Resistance to Fatigue

Fatigue of bituminous mixtures is an important factor in pavement design [17]. Fatigue cracking is considered as one of the major types of distress. It is defined as the reduction of the strength of a material subjected to cyclic loading compared to the strength under non-cyclic loading [28]. Predictive models and laboratory tests are used for fatigue testing [17].

Tests to understand the behavior are destructive, and time-consuming [17]. Shell prescribed equations for constant stresses and strains using the fatigue line of several samples from different types of mixes, bitumen, and testing conditions. [29]. The equations can be translated into a nomograph, however, the accuracy of these equations is 40% for 90% of the constant stresses and 50% for the constant strains [17].

In the Netherlands, fatigue life is determined by bending, direct, and indirect tensile tests [28]. In a four-point bending test, the failure point N_f is defined as the number of cycles required to reach half of the initial stiffness. Fatigue resistance is expressed by the strain level achieved after one million load repetitions, and it is indicated as ε_6 $[\frac{\mu m}{m}]$ [28]. The results from tests following the NEN-EN 12697-24 - method D guidelines were used in this research.

2.4.2 Resistance to Permanent Deformation

Permanent deformation is characterized by the non-reversible deformation under the wheel path, also known as ruts[30]. Rutting is a major form of distress that can cause hydro planning by water accumulation [30]. These deformations are accumulated incrementally over the pavement's life.

Tests to determine the resistance to permanent deformation consist of applying a constant static confining pressure and measuring the axial deformation of a cylindrical specimen under cyclic axial loading [31]. The resistance is expressed as creep rate f_c $[\frac{\mu\varepsilon}{cycle}.10^6]$, which can be measured from the slope of the linear part of the creep curve, also known as turning point [31]. The creep curve derives from the cumulative strain of the test specimen after n loading cycles.

$$\varepsilon_n = A_1 + B_1 \cdot n \tag{2.2}$$

Where:

ε_n: Cumulative axial strain of the test specimen after n loading cycles, in percent.

n: Number of loading cycles.

A_1, B_1: Regression constants.

f_c is then derived from the (quasi)linear part of the curve:

$$f_c = B_1 \cdot 10^4 \tag{2.3}$$

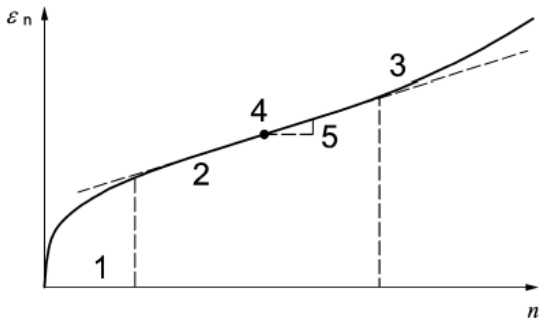


Figure 2.4: Creep curve example [31]

Where:

1: Stage 1 of creep curve.

2: Stage 2 of creep curve.

3: Stage 3 of creep curve.

4: Turning point.

5: Creep rate f_c

The application of a friction reduction system between the plates that perform the axial compression are advised for this test [31]. The friction reduction system is necessary to reduce the friction stresses caused between the plate and the sample and they can have considerable influence on test data [32, 14]. The results from tests following the NEN-EN 12697-25 - method B guidelines were used in this research.

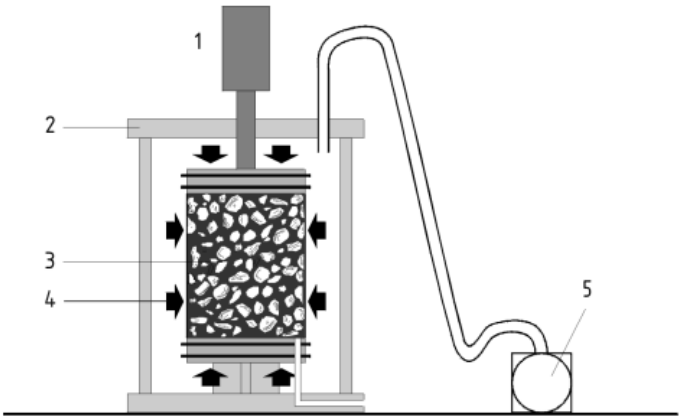


Figure 2.5: Testing apparatus [31]

Where:

1: Actuator for dynamic pressure.

2: Pressure cell.

3: Sealed test specimen.

4: Confining pressure.

5: Compressor.

2.4.3 Water Sensitivity

Pavements are often exposed to traffic and environmental loading combinations that could damage and reduce their performance [15]. Moisture damage is considered as one of the major causes of pavement distress for hot mix asphalt [33]. When the binding strength between bitumen and aggregates is reduced, the susceptibility of stripping is increased [33].

Water sensitivity can be determined by the ITS of cylindrical specimens [34]. The specimen is divided into two identical samples that are subjected to different conditions. One specimen is maintained dry while the other is saturated under elevated conditioning temperature [35]. After conditioning, the ratio between the ITS defines the indirect tensile ratio (ITSR), expressed as percentage [35].

The ITS test is performed by placing the subset into a compression machine and loading with a constant speed, diametrically along the direction of the cylinder axis until it breaks [34]. The tensile strength, expressed in MPa, is calculated from the peak load applied and dimensions of the sample [34]. This research uses the results from tests following the NEN-EN 12697-23 guidelines to fit the indirect tensile strength, and an alternative to assess water sensitivity is suggested.

$$ITS = \frac{2P}{\pi \cdot D \cdot H}.1000 \tag{2.4}$$

Where:

P: Peak load, expressed in Newtowns (N).

D: Diameter of the sample.

H: Height of the sample.

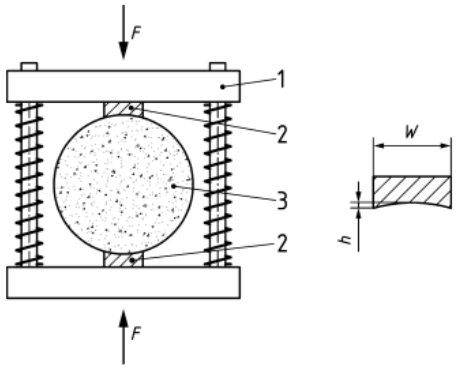


Figure 2.6: Testing apparatus [34]

Where:

1: Testing head.

2: Loading strips.

3: Specimen.

W: Width of the loading strip.

h: Maximum height difference at the curved side of the loading strip.

F: Load.

The functional properties of asphalt mixtures relate to the functional properties of the binder. In Section 2.2, the tests performed on the bitumen were specified.

2.5 Bitumen functional properties

2.5.1 Penetration test

The consistency of asphalt binders can be determined by penetration test [36]. This test does not measure the fundamental properties or control the loading rate [37]. The penetration test measures the vertical penetration depth of a needle into a sample of bituminous material under specific conditions of temperature, load, and loading duration [37]. For penetrations smaller than $330 \cdot 10^{-1}$mm, the testing temperature should be of 25°C with the loading of a 100g needle for 5s. For tests that expect higher penetration values, the testing temperature should be 15°C [36]. The results from tests following the NEN-EN-1426 guidelines were used in this research.

2.5.2 Softening temperature test

The softening temperature of the bituminous binders can be determined by Ring and Ball test [38]. The test is performed by placing two horizontal discs filled with binder inside a liquid bath with steel balls laying on top of the binder [38]. By heating the liquid at a constant rate, the binder will start to deform. The softening point is the mean value in which both balls have fallen a distance of 25mm [38]. The penetration test as referred in the previous subsection and ring and ball tests can be used to determine the Penetration Index (PI), which are inputs for Van Der Poel and Shell nomographs for stiffness determination [17]. In the research, results from tests following the NEN-EN-1427 guidelines were used in this research.

2.5.3 Dynamic shear rheometer

Dynamic shear rheometer (DSR) is used to measure the dynamic visco-elastic properties of asphalt binders [37]. From this test, it is possible to determine the dynamic shear modulus (G^*), and the phase angle (phase lag, δ) [37]. G^* is the ratio of peak shear stress to the peak shear strain in harmonic sinusoidal oscillation, and the δ is the phase difference between the stress and the strain under the same oscillation [39]. The test is conducted using a shear rheometer which is fitted with parallel plates with a constant gap [39]. The diameter of the plate depends on the binder's stiffness [39]. Binders with stiffness ranging from 1kPa to 100kPa require plates of 25mm diameter with a 1mm gap. From 100kPa to 10MPa, the diameter should be 8mm with a 2mm gap. [39]. The test was conducted with temperatures ranging from -10 to 60°C, and the master curve was built at a reference temperature of 20°C and frequency of 10 rad/s [14]. In this research, results from tests following the NEN-EN-14770 guidelines were used in this research.

The properties of the components present in the mixture also impact the performance. The following section provides an overview of the mixing composition and its impacts in its lifetime.

2.6 Mixing composition and properties

2.6.1 Bituminous binder

The bituminous binder is derived from the residuals of distilled crude oil [37]. They are composed of a large number of hydrocarbon molecules [40]. The assessment of the properties is done by the methods described in Section 2.5. The following subsections describe the effect of bitumen content and properties on mixture performance.

2.6.1.1 Bitumen content

The binder content is defined based on the durability and stability of the mix [17]. For continuous graded mixes, it is necessary to find the balance between stability and flexibility [41]. More binder will help with the flexibility and less binder with the stability. The use of high bitumen content can lead to bleeding, in which the material percolates and creates a film on top of the layer [17]. On the other hand, with less binder, the susceptibility to rutting reduces, but the material becomes more brittle and prone to crack [4].

2.6.1.2 Bitumen properties

For rutting prevention, it is desirable to increase the binder's stiffness and decrease the phase angle to decrease the viscous dissipation [37]. Increasing the stiffness of the binder also reduces fatigue cracking for layers thicker than 125mm. Additionally, the increase of the stiffness mixture is related to the increase of the stiffness of the binder [37].

Bitumen aging results in changes in the properties over time [14]. Aging is a complex phenomenon caused by oxidation and volatilization [37]. Oxidation changes the structure of the molecules leading to more brittle behavior [37]. Tests show that the stiffness of an asphalt mixture increases with aging time over all of the frequency ranges leading to higher cracking potential [42]. This cracking potential can lead to space for water to percolate in the binder-aggregate interface, leading to water damage [43].

2.6.2 Reclaimed asphalt

The recycling of hot mix asphalt (HMA) is mainly the incorporation of reclaimed asphalt pavement (RAP) [37]. The reuse of RAP in new mixtures reduces the need for raw aggregates and fresh binders in the construction of new structures [44].

The use of RAP was believed to be limit the ITSR and fatigue resistance [4]. However, results from tests often indicate improvement on ITSR for a high reuse percentage. On the other hand, research has shown that high percentages of RAP lead to a lower fatigue resistance [4]. Additionally, tests indicated that the increasing RAP percentage improves the complex modulus of mixtures and resistance to permanent deformation [45].

Tests indicate that the aged bitumen in RA increases the softening point and viscosity and decreases the penetration of the fresh binder [44]. Moreover, the viscosity can also increase in proportion to the amount of RAP added to the mix [44]. With the increasing viscosity, the workability of the mix is affected [44]. The viscosity reduction can be achieved by mixing the reclaimed binder with a softer fresh binder [44].

2.6.3 Aggregates

Asphalt mixtures consist of bituminous binder, mineral aggregates, and air voids [46]. They constitute 70 to 85% of the mixture's weight [37]. The mechanical properties of aggregates, stone, and sand, relate to the resistance to degradation, polishing, impact, or loading [37]. Even though this research does not consider these properties, it is important to highlight them for a deeper understanding of the results.

Aggregates with higher stiffness are desired for most applications [37]. The angularity impacts the rutting resistance of the mix because of the internal friction and the chemical properties affect the stripping and raveling on the presence of moisture [37]. The aggregate's quality also affects the mixing and compaction stages of the design. Weak aggregates can fail during these stages, reducing the project's durability. The breakage can also occur due to traffic loads, especially on gap-graded pavements that rely on stone-on-stone contact [37].

2.6.4 Fillers

The filler is part of the aggregate percentage and can be characterized as minerals passing through a sieve 0.063mm [46]. The aggregates, larger than filler, are not only coated by the binder but by the mastic [46]. Mastic is the mixture between the binder and the filler, which has great significance in the performance of flexible pavements [46].

In the study by Huang et al, the indirect tensile strength of mixtures was found to increase when the filler content was increased [46]. The increase in ITS could be because the strength of the filler is higher than the strength of the binder, and by increasing the filler content and reducing the binder content, the strength goes up. The water sensitivity ratio, on the other hand, decreased. The reduction in ITSR could be explained by the reduced binder content with the increased amount of filler [46].

Resistance to permanent deformation decreased when the filler content was increased from 2% to 5% [46]. However, when it increased to 10%, the rut depths were smaller when compared to the smaller percentages, improving the resistance to permanent deformation. The reason could be because of the increase in stiffness caused by the filler content [46].

Stiffness and fatigue life could also increase as the filler percentage increases [47]. Stiffer mastic results in higher mixture stiffnesses, and the extra filler particles interrupt the crack propagation [47].

The possibility of using data-driven approaches to predict the functional properties of asphalt mixtures is an attractive alternative to diving into the applicability of models and understanding their functionalities. The following section aims to provide a brief explanation of data-driven approaches and their branches.

2.7 Data-driven approaches

The Cambridge Dictionary defines data-driven as something based on collected information [48]. Machine learning and statistical models base the decision rules on the collected data [49]. Machine learning is a subset of artificial intelligence where machines improve in a specific task with experience [50]. It is divided into three subsets [51]:

- Reinforcement learning
- Unsupervised learning
- Supervised learning

Reinforcement learning, the model that learns from a series of rewards or punishments [51]. The concept is to understand what caused the reward or punishment aiming to maximize the total reward. In unsupervised learning, the model learns hidden patterns in the data without labeling. A typical unsupervised learning technique is clustering [51]. In supervised learning, the model learns patterns in labeled data to make predictions [52]. The predictions can be:

- Classification: categorizes observations in groups
- Regression: forecasts continuous values

This research applies regression models to predict the functional properties of asphalt mixtures described in Chapter 1. The techniques applied are further explained in the following chapter.

2.8 Conclusion

This chapter presented the introduction to the research, along with an investigation of the literature providing the background for the parameters being studied, the influence of components,

and different tools applied for the determination of the functional properties of asphalt mixtures. The finding highlighted the applicability of tools and the impact that different material properties have on the mixture. The following chapter describes the methods and tools utilized in this research.

3

Methodology

The methodology of this research is structured in four main parts. (i) The first part consists of data preprocessing and feature engineering. (ii) In the second part, regression analysis, optimization, and cross-validation is present. (iii) The third part focused on sensitivity analysis. (iv) The last part presents the tools used for the regression analysis. The first three parts of the methodology follow the structure shown in Figure 1.1.

3.1 Data preprocessing and feature engineering

Data preprocessing and feature engineering are major steps in statistical and machine learning models [53]. The data was used in this research gathered in raw form, and techniques were applied to increase the quality of information given to the models. The following techniques were applied:

- Data cleaning
- Data reduction
- Data integration
- Data transformation
- Feature creation
- Feature selection

Data cleaning consists of removing inconsistencies, filling in missing values, and smoothing noisy data [53]. Data reduction filters the necessary data from the mater dataset [53]. Data cleaning and reduction were covered in a single step. Data integration was used to avoid conflicts between

variables [53].

Data transformation is a step that relates to both data preprocessing and feature engineering [54]. In data transformation, the datasets are scaled and encoded [53]. Scaling of the data is done to avoid large deviations [53]. In addition, in the dataset, there are numerical and categorical variables, which nominal data that must be converted into numerical values.

In feature creation process, variables that better translate to the model were created [55]. In feature selection, the final dependent and independent variables were selected. It is essential to highlight that these processes are iterative, and no specific order was followed [54]. During the modeling process, the preprocessing and engineering of the data were performed multiple times. The following subsection focuses on explaining the above mentioned steps.

3.1.1 Data cleaning and reduction

The inputs collected from the master dataset consisted of type tests, material composition, material tests, sample codes, mixing and compacting equipment, and age of sample. The raw data extraction resulted in 79 features (see Appendix B.1).

Features with a large number of missing values and variables that can not be determined in advance were removed. It is necessary to focus on parameters that are related to most of the properties. The initially removed items were parameters that were simply for the organization and bookkeeping of the samples, the regression coefficients from the tests were used to determine the dependent variables, and the extracted and reference composition.

The data, even though gathered in the same master dataset, come from different tests and places [15]. The different data sources can lead to conflicting conclusions [53]. The data integration process that was followed in the research is explained in the next subsections.

3.1.2 Data integration

As explained in Section 2.2, the data come from various laboratory tests of asphalt concrete mixtures, and their components [14]. Parameters in the analysis can become redundant by translating information from other variables. In this step, the data incorporating information from other parameters, was removed. Additionally, two correlation analyses were performed. At first, the correlation between bitumen properties. In the second step, the correlation between the density properties was obtained.

Before transforming, the data was split into the training set and testing set. The splitting methods and transformation techniques are explained in the following subsections.

3.1.3 Splitting

The process of fitting a regression requires splitting the data into two sets [56]. The concept is to develop a hypothesis using the training set and test the hypothesis in the testing set [56].

The data was split into 80% for training and 20%, as commonly adopted in statistical analysis. For this step, the distribution of the entire dataset was evaluated to make sure that the splits respected the original distribution. The same dataset was used for the statistical and machine learning to make reliable conclusions.

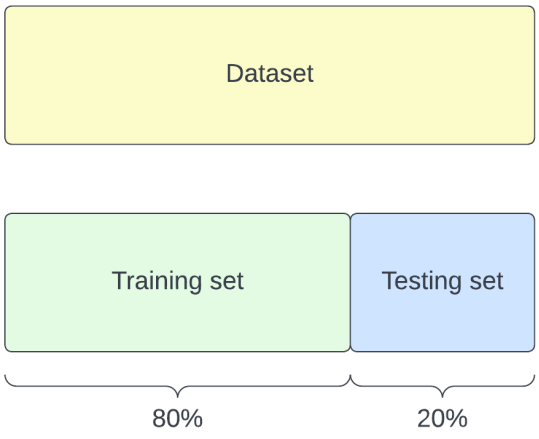


Figure 3.1: Splitting scheme

The split dataset was then used for feature transformation. In the following subsection, a brief explanation of the transformations is given.

3.1.4 Data transformation

The transformation of the data was done in two main steps. The nominal data were converted into numerical values [53]. Additionally, the numerical data was scaled as some models require them to be scaled.

3.1.4.1 Categorical data

Some features were categorical, which means that they were labels and not numbers [50]. For the analysis, it is required to transform them into numerical variables before fitting the model. The most common approach is *One-Hot-Encoding* [22]. It transforms a single variable with n observations and d distinct values into a binary value d with n observations, also known as *dummy variables* [57]. An example of encoding is presented in Table 3.1.

Table 3.1: One-Hot-Encoding example

Mixing set-up
Planetary mixer
Forced action mixer
Asphalt plant

Planetary mixer	Forced action mixer	Asphalt plant
1	0	0
0	1	0
0	0	1

Linear regressions require that one item from the encoded data be deleted [58]. However, the values are still considered in the intercept of the linear regression equation [58]. One of the dummy variables is deleted from the feature list to avoid multicollinearity between the variables [58]. Table 3.1 represents the *One-Hot-Encoding* for random forests and support vector machines since they are not affected by multicollinearity. It was obsorved that removing the asphalt plant column exemplifies *One-Hot-Encoding* for linear regression.

Table 3.2: One-Hot-Encoding MLR example

Mixing set-up
Planetary mixer
Forced action mixer
Asphalt plant

Planetary mixer	Forced action mixer
1	0
0	1
0	0

The transformed categorical features create a column for each unique value in the dataset [59]. The newly created columns reduce the model's performance because it occupies more space. Moreover, more data is required to distinguish the difference between them. The solution was the aggregation of these unique values into reduced groups. A description of reduced data can be found in Appendix B.

Different models applied in this research use different encoding techniques [60]. Target-based encoding was applied, in which the data is replaced with the average target value of the corresponding category [60]. The issue of this method is target leakage, in which information about the value to be predicted is leaked to the model, and it tends to over-fit [60]. The solution for target leakage can be achieved by using the average target value of the occurrences of a specific category before the one being encoded [60]. This approach leads to a high variance in the initial values [61, 60]. Multiple permutations are done to solve the high variance, and the final value is the average of all permutations.

In addition to encoding, some models require numerical features to be scaled to avoid high variances [53]. It is noted that the categorical features converted into numerical features do not require scaling. The following subsection provides strategies used for scaling the numerical data.

3.1.4.2 Feature scaling

Some models generally struggle to deal with data with different ranges, and scaling becomes a necessary step [62]. For instance, several components of a learning algorithm's objective function assume that all characteristics are centered around 0 and have a variance distributed in the same order [62]. A feature may dominate the objective function and prevent the estimator from successfully inferring from other features as expected if its variance is orders of magnitude greater than that of other features.

In this research, linear regression and support vector machines were explored models that require data scaling. The approaches for this step were different for machine learning and statistical analysis. The statistical analysis projected the minimum value as -1 and the maximum as 1 to be uniform, making it possible to compare the coefficients directly. For a continuous variable X, with minimum X_{min} and maximum X_{max}, each value $x \in [X_{min}, X_{max}]$, the scaled equivalent x_{scaled} is calculated using the following formula:

$$x_{scaled} = \frac{x - X_{min} - 0.5 \cdot range}{0.5 \cdot range} \tag{3.1}$$

Where:

range: $X_{max} - X_{min}$.

In machine learning, only the support vector machines required scaling [63]. Standardization was applied using the following formula:

$$X_{norm} = \frac{X - u}{s} \tag{3.2}$$

Where:

u: mean value of the sample.

s: standard deviation of the sample.

Features that contained relevant meaning for the analysis were created, and the final set of variables was selected. Defining the relevant features for the analysis is explained in the following subsection.

3.2 Feature creation and selection

3.2.1 Fatigue resistance extrapolation

Fatigue resistance is measured based on the fatigue line fit from 18 samples and to generate more data points, the available fatigue lines were extrapolated. From the fatigue life plot, a horizontal line was drawn at the $\log 10^6$ cycles point. Parallel lines, respecting the slope of the original line, were drawn from the observations until they intercepted the horizontal line. The strain level for these individual points at $\log 10^6$ cycles was named $\varepsilon_6 Individual$.

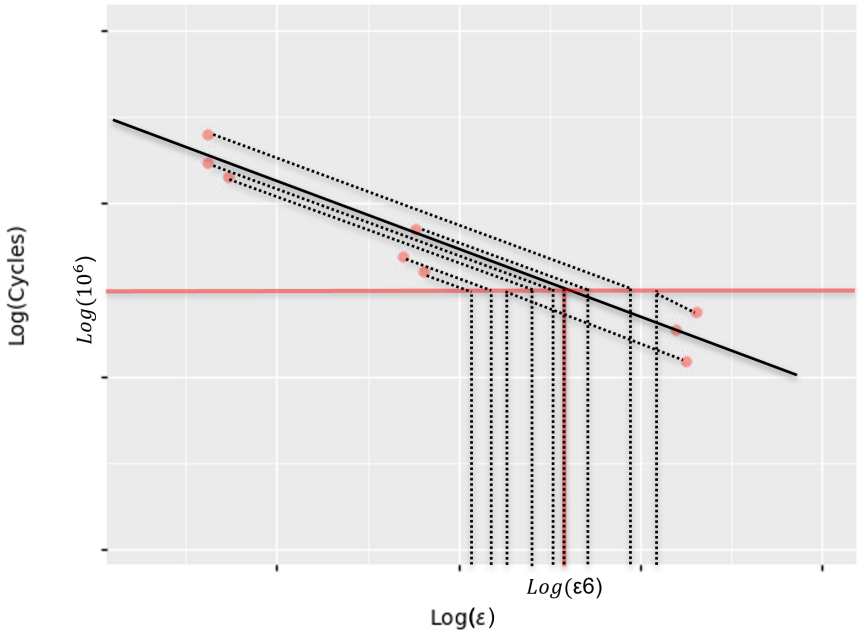


Figure 3.2: Fatigue line extrapolation example

3.2.2 Target mixing composition

The target mass composition of aggregates was measured in percentage passing through the sieves. The percentages were converted into the percentage retained in the sieves. Transforming from passing to retained percentages allowed the creation of stone and sand percentage variables. The two parameters were created to compare the model's performance using group and individual composition parameters. In addition, the target bitumen percentage was not incorporated into the dataset. To incorporate, the total target mass percentage summed with the bitumen percentage should result in 100%. Thus, the target bitumen percentage was calculated by subtracting the total target mass percentage from 100%.

Originally, asphalt mixtures were designed using volumetric properties and later converted to mass properties when being produced [32]. The mass percentages were converted into volume percentages to analyze the behavior of the models. The composition did not specify all materials, and the densities for the different components had to be assumed constant. Table 3.3 shows

the assumed densities for each material. Models considering mass and volume percentages were created to analyze their impact on performance.

Table 3.3: Assumed densities

Component	**Density** $[\frac{kg}{m^3}]$
Stone	2700
Sand	2700
Filler	2500
Bitumen	1030

3.2.3 Compaction degree

The last parameter to be created was compaction degree. The compaction effort is considered to be of major importance because it defined the air void percentage of the mixture [23]. Compaction degree can be determined by the following equation:

$$CompactionDegree = \frac{Density}{TargetDensity} * 100 \tag{3.3}$$

In the next stage, the statistical and machine learning models were built for the regression analysis. The models were later optimized and cross-validated. The following sections provide an overview of regression analysis and the optimization and cross-validation process.

3.3 Regression analysis

Regression is a common statistical technique used to model systems defining a mathematical relationship between a response variable (y) and explanatory variables (x) [50]. Regression analysis can have different benefits [64]. They help in:

- determining if independent variables have a significant relationship with the dependent variable
- estimating the effects of the independent variable on the dependent variable
- making predictions

The regression analysis consists of fitting the line into a group of observations that results in the minimum sum of residuals [64].

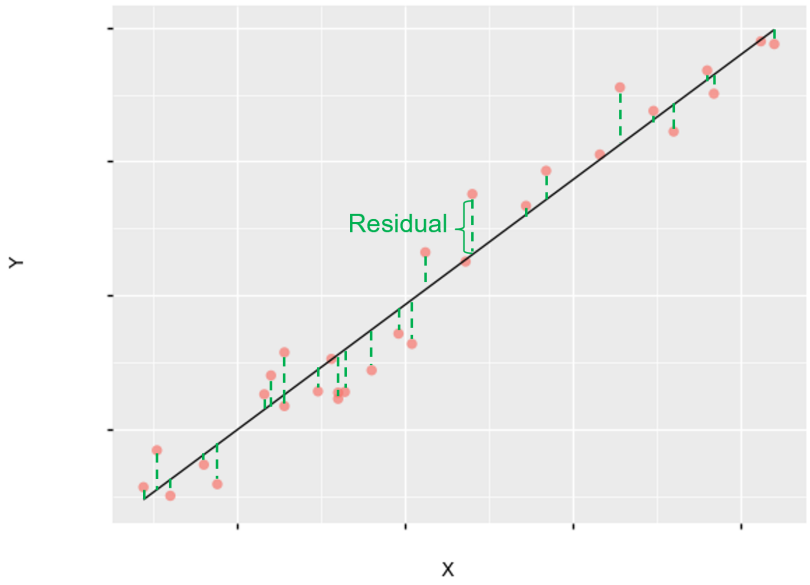
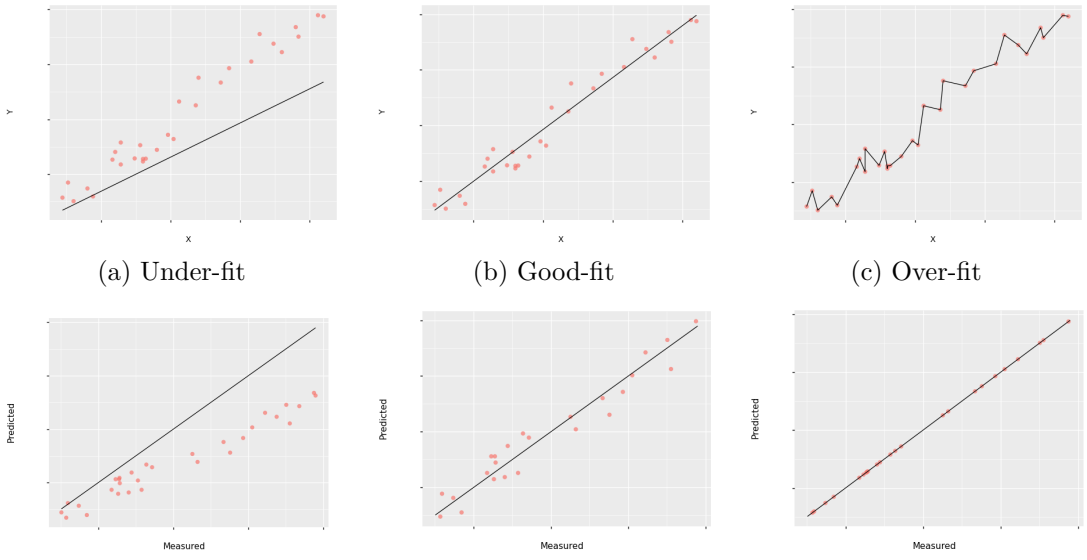


Figure 3.3: Fitting and residuals example

The steps to perform a regression analysis consists of defining the data, creating a mathematical model, evaluating the goodness of the fit, and a fitting strategy [22]. The NL-LAB data was used in this research, as highlighted in Chapter 1. The objective was that the mathematical model selected should balance between over-fitting and under-fitting [50]. In under-fitting, the algorithm does not gather sufficient information to accomplish the learning task [65]. It is avoided by applying a mathematical model that matches the complexity level of the data and selecting a sufficient number of features for the analysis [65]. On the other hand, over-fitting means that the algorithm extracted more information than necessary from the data. Possible techniques to prevent over-fitting are splitting the data into training sets, testing sets, and applying cross-validation [66].



(a) Under-fit (b) Good-fit (c) Over-fit

(d) Under-fit prediction (e) Good-fit prediction (f) Over-fit prediction

Figure 3.4: Fitting Examples

The balance between over-fitting and under-fitting is commonly measured by the coefficient of determination [67]. The coefficient is a measure of the amount that the regression can explain the target value [67]. The values can range from $-\infty$ to 1 [22]. The results provided by the mathematical model should balance between high and low [22]. The values aimed in this research range from 70% to 98%. R^2 is expressed as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \cdot 100\% \tag{3.4}$$

$$SS_{res} = \sum_{n=1}^{N} (y_n - f_n)^2 \tag{3.5}$$

$$SS_{tot} = \sum_{n=1}^{N} (y_n - \overline{y})^2 \tag{3.6}$$

Where:

SS_{res}: sum of squared residuals.

SS_{tot}: sum of squares.

y_n: response variable.

f_n: predicted value.

$\overline{y}$: mean value of response variable.

The models selected to perform the regression analysis were multiple linear regression (MLR) for the statistical analysis and support vector regressor (SVR), random forests (RF), and gradient boosting (GB) for the machine learning analysis. MLR has been part of pavement engineering for a long time [6]. Showing good results for stiffness prediction [18, 22]. MLR underperformed in predicting the creep rate and ITS. However, it is a common tool, relatively simple to apply and interpret [7]. SVR has shown prediction performance superior to Witczak model [68]. RF had high-performance accuracy for stiffness prediction [69] GB had a high performance for stiffness, creep rate, and ITS prediction [22].

After the models have been built, optimizing and generalizing them is necessary. The optimization process and cross-validation of the model are described in the following section.

3.4 Optimization

Optimizing the models, also known as tuning, is a critical step for any machine learning [70]. Hyperparameters are the parameters used to calibrate a machine learning model or reduce

an objective function [71]. The hyperparameters to be tuned were set beforehand, and the algorithm automatically found the optimum set. Hyperparameter is optimized by using the following equation [71]:

$$x^* = arg \min_{x \in \chi} f(x) \tag{3.7}$$

where;

f(x): the objective function to be minimized.

x*: set of hyper-parameters that yields the minimum value of f(x).

There are different methods to tune the model, such as, manual, Grid Search, Random Search, and Bayesian Optimization [71]. The application of Grid Search and Random Search has been found to be more efficient than the manual process. However, it is computationally demanding since the model will iterate through the list of parameters and test all possible combinations that minimize the previous function.

Bayesian Optimization concept, on the other hand, has shown superior results than random and grid search [72]. It keeps track of past evaluations to decide which one will be done next. It uses a probabilistic method to map the hyper-parameters to the probability of a score on the objective function. This method is called the surrogate model, and it is represented by the probability of a score y to the given hyper-parameters x (p(y|x)).

The application of such a concept consists of a Sequential model-based optimization (SMBO) method [73]. In every trial, the model updates based on the previous trial. The surrogate model used in this research consists of a Tree Parzen Estimator (TPE) with the Expected Improvement criteria (EI) [73].

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y) p(y|x) dy \tag{3.8}$$

Where:

y*: threshold value of objective function.

x: set of hyper-parameters.

y: the actual value of the objective function using x.

p(y|x): surrogate probability of y given x.

In this function, if the probability is zero, the result will not have any improvement [73]. In order to have improvements, the output of the integral must be positive. TPE uses a Bayes rule to describe p(y|x).

$$p(y|x) = \frac{p(x|y) \cdot p(y)}{p(x)} \tag{3.9}$$

$p(x|y)$ can be defined by the following:

$$p(x|y) = \begin{cases} l(x) & if\ y < y^* \\ g(x) & if\ y > y^* \end{cases} \tag{3.10}$$

Where:

l(x): density of observation below threshold.

g(x): density of observation above threshold.

The algorithm defines the threshold as a quantile (γ) of the observed y values [73]. By substituting into the EI function, it yields:

$$EI_{y^*}(x) = \frac{\gamma y^* l(x) - l(x)\int_{-\infty}^{y^*} p(y)dy}{\gamma l(x) + (1-\gamma)g(x)} \propto (\gamma + \frac{g(x)}{l(x)}(1-\gamma)^{-1}) \tag{3.11}$$

The last part of the equation means that to maximize the EI function, it is necessary to collect the high probability points in l(x) and low in g(x) [73]. In summary, it makes a probability model of the objective function and decides which are the most promising hyper-parameters to test in the real objective function [73].

In addition to the optimization of hyper-parameters, loss functions play an essential role in the performance improvement of machine learning models [74]. The next subsection highlight some of the loss functions that were applied in the analysis. It is important to note that different loss functions were only applied at the final stage of modeling on the selected model to perform a sensitivity analysis.

3.4.1 Loss function

During the training and optimization process, the models need a reference function to minimize the error [74]. The loss function selection affects the model's accuracy, and efficiency [74]. All models in the analysis were trained and optimized using the root mean squared error (RMSE). Later, the models that yielded the highest performance were trained in two extra loss functions to understand how it would impact the overall performance. Quantile loss and expectile loss were also tested.

The RMSE resembles the Euclidean distance [75], meaning that it can be defined as a distance from the prediction to the actual value. By dividing by a factor N, the Euclidian distance is re-scaled, allowing the estimation of the standard deviation.

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(y_n - f_n)^2}{N}} \tag{3.12}$$

Where:

n: variable n.

y_n: observation value.

f_n: prediction value.

N: number of data points.

Quantile loss can provide a prediction interval, resulting in the range of prediction instead of a single value [74]. The function adjusts the asymmetric weights, $w_{n,\alpha}$, based on the selected quantile, α [74]. α values range from 0 to 1 and are used to penalize the prediction [76]. For values of α above 0.5, the over predictions are penalized. Below 0.5, the under predictions are penalized, and 0.5 penalizes equally.

$$Quantile = \sum_{n=1}^{N} w_{n,\alpha}|y_n - f_{n,\alpha}| \tag{3.13}$$

Where:

$$w_{n,\alpha} = \begin{cases} 1-\alpha, & for\ y_n < f_{n,\alpha} \\ \alpha & for\ y_n \geq f_{n,\alpha} \end{cases} \tag{3.14}$$

Expectile uses the same principle as quantiles. However, the distance ($|y_n - f_{n,\alpha}|$) is a quadratic term leading to asymmetric least squares [76].

$$Expectile = \sum_{n=1}^{N} w_{n,\tau}(y_n - f_{n,\tau})^2 \tag{3.15}$$

Both Quantile and Expectile had α of 0.5.

The models had to be generalized together with the optimization process because during the split process, the distribution of data was controlled. The generalization was made by using a 10-fold cross-validation method. An explanation of the cross-validation process can be found in the following section.

3.5 Cross-Validation

Cross-validation is the commonly used approach to generalize the models [77]. The training and testing data distribution respected the original dataset's distribution, as explained in Section 3.1. However, controlling the distributions is not always possible, and cross-validation is required. The dataset is divided into smaller samples that naturally will have different distributions. The models are trained in these smaller samples, and the results are averaged [77].

In k-fold cross-validation, the data is split into k-equal samples [77]. The model is trained in k-1 samples, and the remaining set is used for validation. This process is repeated by exchanging the validation set for one of the training sets. The process is repeated until all k-samples have been used for training and testing [77].

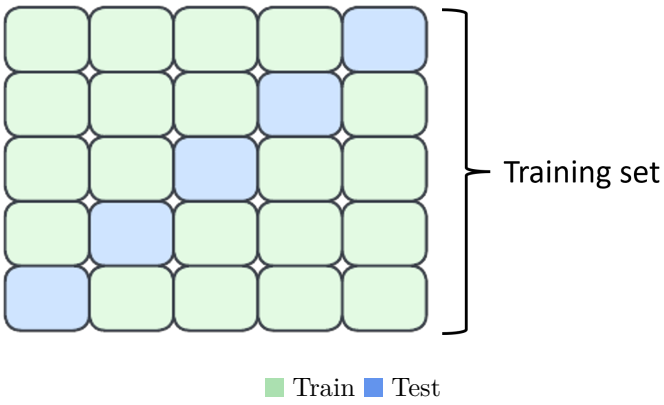


Figure 3.5: 5-fold Cross-validation Example

The cross-validation of the models finalized the last step of optimization. Next, a single model for each functional parameter was selected, and a sensitivity analysis was performed. The following section gives an overview of sensitivity analysis.

3.6 Model sensitivity

Sensitivity analysis identifies the important parameters in the analysis and their impact [50]. It evaluates the changes in the result based on the change of inputs [50]. Linear regressions, for example, are fairly simple to understand and interpret, and many times are selected because of such ease. For more complex models, more robust methods to interpret the results are necessary [7]. SHAP values can be used for interpreting more complex models [7]. It stands for Shapley Additive exPlanations and is based on the game theory [7].

Lloyd Shapley introduced Shapley value in his publication [78]. The goal is to understand the contribution of a specific member of a coalition to produce value. The contribution of a member

of the coalition is calculated by comparing the differences between the values of this coalition with and without the member. The differences yield the marginal contribution of the member [7]. The Shapley value will be the mean contribution of the member considering all possible combinations in which it is included.

Considering a model f(x) that is represented by the input x, has to be explained by an explanatory model g($x^{'}$) represented by simplified inputs $x^{'}$, which is the transformation of a feature vector into a discrete binary vector [7]. It is explained in the following notation:

$$f(x) = g(x^{'}) = \phi_0 + \sum_{i=1}^{M} \phi_i x_i^{'} \tag{3.16}$$

where:

ϕ_0: null output of the model.

ϕ_i: feature effect (attribution).

Three properties have to be respected [7]:

- local accuracy: states that if the input x $\approx x^{'}$ then $g(x^{'}) \approx f(x)$
- missingness: states that if a feature is excluded from the model, its attribution must be zero ($x_i^{'} = 0 \longrightarrow \phi_i = 0$)
- consistency: states that if a simplified input's contribution changes or stays the same, the input's attribution should not do the opposite

The SHAP values can be represented in several types of plots [79]. In this research, summary plots and scatter plots are used. In the summary plot, Figure 3.6, the features listed on the left-hand side are ranked by their importance [79]. The mean absolute value of the SHAP values for each observation defines the order. In the second, on the right-hand side is a color bar representing the features' real value. Blue color stands for the minimum value of the feature in the dataset, red is the maximum, and the gradient changes while the values progress from the minimum to the maximum. Grey values represent categorical features, which have been converted from numerical values back to text values. The values for categorical features can be expanded into a summary plot for each feature. The x-axis represents the impact of the data points on the model. The more it shifts to the right, the higher the positive impact on the prediction, and the more to the left, the higher the negative impact. Points located in the middle represent that the individual contribution was zero.

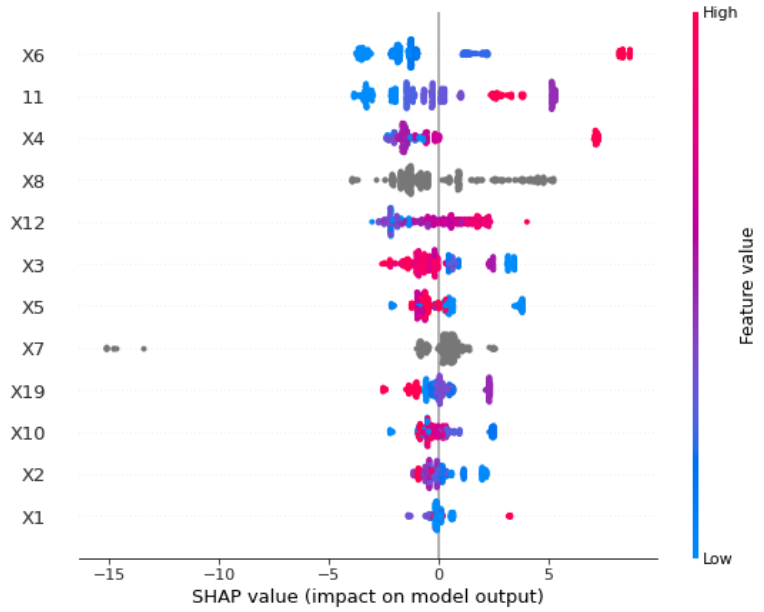


Figure 3.6: SHAP summary plot example

The values shown in summary plots can be translated into scatter plots. In Figure 3.7, the x-axis has the feature values, explained as a color gradient in the summary plot. The y-axis is the impact of the data point in the model, the same represented by the x-axis in the summary plot.

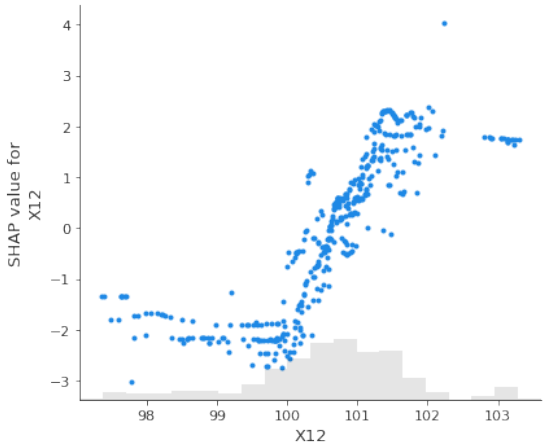


Figure 3.7: SHAP scatter plot example [79]

The following section gives an overview of the background and mathematical formulation of the regression tools used. The machine learning tools that were used has been divided into two groups, support vector machines and ensembles.

3.7 Tools

3.7.1 Multiple linear regression

The linear dependency between explanatory and response variables can be modeled by linear regression [49]. Models with a single explanatory variable are called simple linear regressions, while models with more than one are multiple linear regressions. However, MLR requires some assumptions to be met[22]:

- Normal distribution of residuals
- No or little multicollinearity
- Homoscedasticity

A generic formulation of The MLR is formulated as the following [67]:

$$Y = \beta_0 + \sum_{i=1}^{D} \beta_i \cdot X_i + \varepsilon$$

Where:

β_0: intercept in the y-axis.

β_i: slope of the line

ε: model uncertainty.

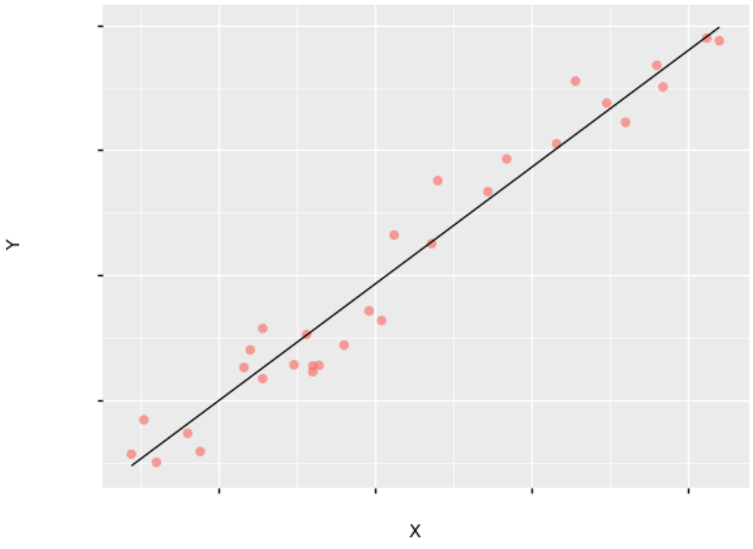


Figure 3.8: Linear regression example

Ordinary least squares are commonly used for the model's estimation (OLS) of a regression analysis [64]. OLS is the distance between the observation and the fit line. These distances are squared to prevent negative values from canceling positive values, and the line is fit to minimize the sum OLS.

Support vector machines can have similar outcome to MLR depending on the type of approach being used [80]. The following subsection explains support vector machines and the different approaches.

3.7.2 Support Vector Machine

Support vector machines (SVMs) are supervised learning algorithms that can be used for classification, regression, and outlier detection [80]. It is based on Vapnik-Chernonenkis (VC) theory. SVMs are originally structured for classification problems known as support vector classification (SVC). Regression models are a generalization of SVC and are known as support vector regression (SVR) and can be linear or non-linear depending on the type or kernel used [81].

3.7.2.1 Linear SVR

In linear SVRs, the best fit for the model is not represented by a single regression line, as linear regression, but by two lines offset vertically from the main line [81]. These lines create the shape of a tube around the regression line, named ε-Insensitive Tube [81]. Points within this tube have the errors disregarded, and the ones that fall outside the tube will have the errors calculated until the boundaries, as shown in Figure 3.9. The error is calculated by minimizing the sum of the distances, ξ^*, calculated by the following:

$$\frac{1}{2} \parallel w \parallel + c \sum_{i=1}^{m} (\xi_i + \xi_i^*) \Rightarrow min \tag{3.17}$$

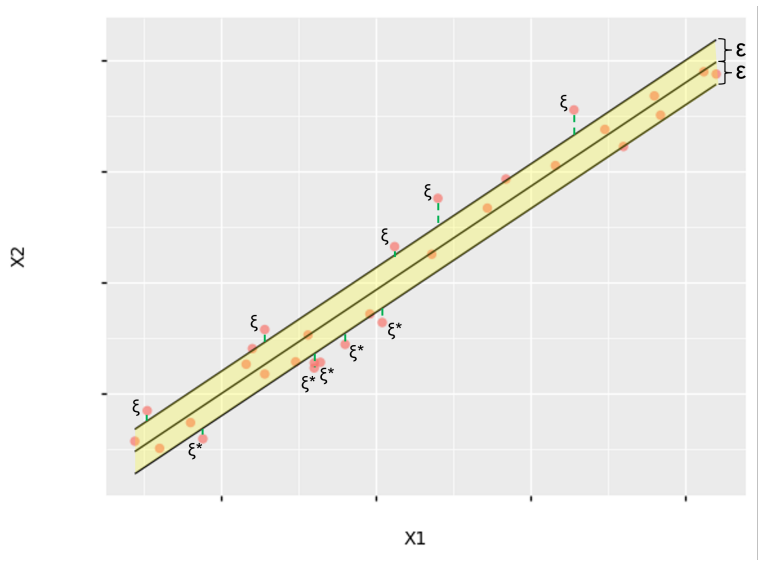


Figure 3.9: Support Vector Regression

The name support vector comes from the slack variables (ξ^*), which are treated as vectors and support the ε-Insensitive Tube, giving more flexibility compared to a linear regression [81].

3.7.2.2 Non-linear SVR

The application of linear or non-linear SVR comes from the method by which the data is divided. Often, a single line does not describe the data and more advanced methods are required [81].

The kernel method is used to make a linear fit in non-linear data. One example of a kernel is the Gaussian radial basis function (RBF) which is described as the following [63, 80]:

$$K(\overrightarrow{x}, \overrightarrow{l^i}) = e^{-\frac{\|\overrightarrow{x} - \overrightarrow{l^i}\|}{2\sigma^2}} \tag{3.18}$$

Where:

K: kernel.

$\overrightarrow{x}$: point in the dataset.

$\overrightarrow{l^i}$: center of the function.

i: different centers.

σ: decides how much curvature we want in a decision boundary.

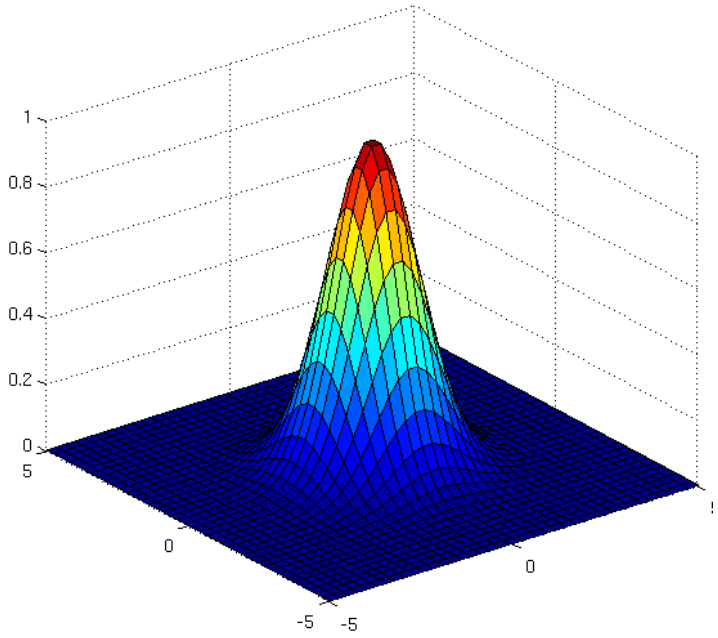


Figure 3.10: RBF function Kernel [82]

The center, in RBF, is the point located in the (0,0,0) coordinate of Figure 3.10 [83]. The vertical axis is the result of the kernel calculation. When there is a large distance between

the data point and the center, the value in which Euler's number is raised is also large, and the equation tends to zero. For small distances, the result tends to be 1. From the kernel equation, the original data points are projected into the 3D-shape of Figure 3.10 and often a linear model is fit. The result of a linear model in three dimensions is a hyperplane where the intersection between the hyperplane and function is the non-linear fit of the original data points [81]. By offsetting this hyperplane by ε upwards and downwards, the non-linear ε-Insensitive Tube is defined. The optimum values for ε and σ are defined through model optimization. The mathematical formulation is expressed as the following [63]:

The training vectors $x_i \in \mathbb{R}^p, i = 1, ..., n$, and $y \in \mathbb{R}^n$ ε -SVR resolves for:

$$\begin{aligned} \min_{w,b,\zeta,\zeta^*} \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\zeta_i + \zeta_i^*) \\ \text{subject to } y_i - w^T \phi(x_i) - b \leq \varepsilon + \zeta_i, \\ w^T \phi(x_i) + b - y_i \leq \varepsilon + \zeta_i^*, \\ \zeta_i, \zeta_i^* \geq 0, i = 1, ..., n \end{aligned} \tag{3.19}$$

Thus, the samples in which the predictions are outside ε are penalized by ζ_i or ζ_i^*, depending on whether their predictions lie above or below the ε tube.

The dual problem is

$$\begin{aligned} \min_{\alpha,\alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \varepsilon e^T (\alpha + \alpha^*) - y^T (\alpha - \alpha^*) \\ \text{subject to } e^T (\alpha - \alpha^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, ..., n \end{aligned} \tag{3.20}$$

The vectors are mapped into a higher dimension space by the kernel function ϕ.

The prediction is:

$$\sum_{i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b. \tag{3.21}$$

The analysis was also performed with decision tree-based ensemble methods. The following subsection gives an overview of ensembles and decision trees. Later, random forests and gradient boost models are explained.

3.7.3 Ensembles

Ensemble models use multiple algorithms to create a better prediction [49]. Random forests (RF) and gradient boosting (GB) algorithms use several decision trees to build the prediction [49]. Decision trees are a non-parametric supervised learning method used for classification and regression [84]. The objective of such trees is to learn decision rules derived from the features in a piecewise constant approximation to build a model that predicts the value of a target variable [84].

A tree structure is shown in Figure 3.11. It is composed of a root node that represents the entire dataset, the internal node represents the conditions that must be met, branches provide the decision rules, "yes/no", and the leaf is the decision [22].

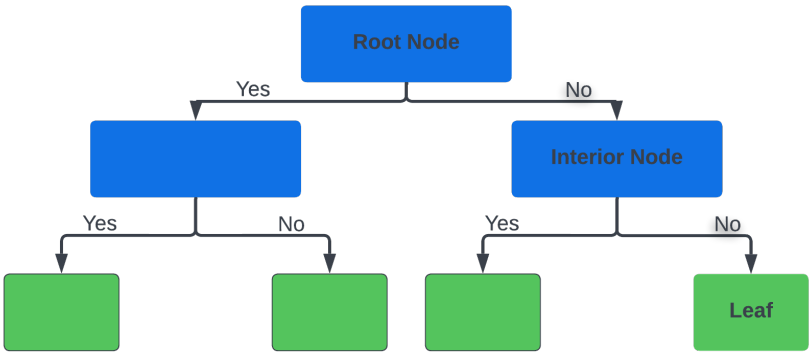


Figure 3.11: Decision Tree

Decision trees tend to over-fit data with many features [84]. To avoid over-fitting, limiting the amount of observation inside the splits is necessary so the model can stop once this threshold is reached. The mathematical formulation is the following [84]:

Given training vectors $x_i \in R^n, i =, ..., |$ and the dependent variable vector $y \in R^l$, a decision tree recursively partitions the dataset, grouping the samples with the same labels or similar target values.

Let Q_m represent the data at node m with n_m samples. Each candidate split $\theta = (j, t_m)$ consisting of a feature j and threshold t_m, partitions the data into $Q_m^{left}(\theta)$ and $Q_m^{right}(\theta)$ subsets

$$\begin{aligned} Q_m^{left}(\theta) &= \{(x, y)|x_j \leq t_m\} \\ Q_m^{right}(\theta) &= Q_m \setminus Q_m^{left}(\theta). \end{aligned} \tag{3.22}$$

A loss function $H()$ is used to compute the quality of a candidate split of node m

$$G(Q_m, \theta) = \frac{n_m^{left}}{n_m} H(Q_m^{left}(\theta)) + \frac{n_m^{right}}{n_m} H(Q_m^{right}(\theta)). \tag{3.23}$$

Selects the parameters that minimize the function until maximum allowed depth is reached

$$\theta^* = \text{argmin}_\theta \, G(Q_m, \theta) \tag{3.24}$$

The RF and GB models required the notion of decision trees. The following subsection aims to provide background information on RF and GB models.

3.7.3.1 Random Forests

Decision trees are estimators that learn simple decision rules [84]. However, when the complexity of such trees starts to increase, the results are a loss of accuracy, and it becomes hard to generalize and add new data [85]. The construction of random forests overcomes these limitations. The name random forests come from using multiple decision trees to predict [85]. The first step to building a model is to construct a bootstrapped dataset [49]. The bootstrap concept is to randomly select data points in the original dataset to generate a new dataset with the same size as the original, in which observations can be repeated [50]. The trees are built by selecting a random subset of features from the bootstrapped dataset and repeating this process multiple times [85]. Building trees from the bootstrapped dataset gives variety to the model since different combinations are integrated into the trees. The final prediction is the average of all target values predicted by the trees.

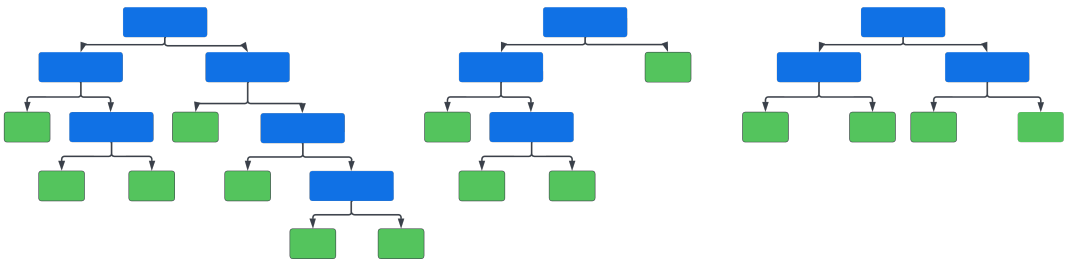


Figure 3.12: Random forests structure

3.7.3.2 Gradient Boosting

The concept of gradient boosting was developed in 1990 by Robert Schapire when answering the question if set weak learners (decision trees) could result in a good predictor [86]. This question was first raised by Kearns and Valiant in 1988 [87].

Gradient boosting algorithms initially build the first tree based on the average of the target values [88]. The residuals between the first prediction and the target become the new value to be predicted. In an iterative process, decision trees are built in series, reducing the error of the previous tree [49]. The final result is the sum of all predictions with the average value. The sum of all predictions could lead to over-fitting. Hence, the predicted residuals are scaled to increment the averaged value in small steps. Figure 3.13 represents the structure of GB, where the 0.1 factor multiplying the tree is the scaling, also known as the learning rate.

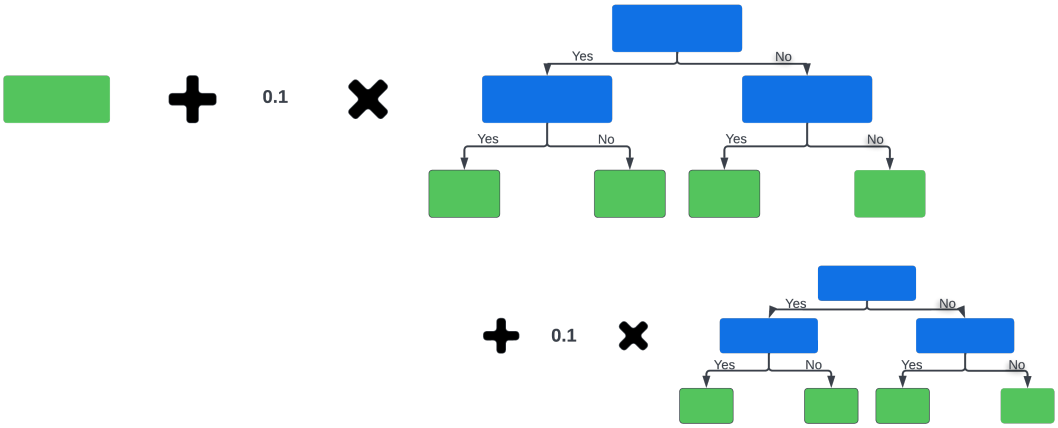


Figure 3.13: Gradient boosting structure

The mathematical formulation is the following [89]:

Gradient boosting regression trees regressors add the predicted $\hat{y}_i$ for x_i in the following form:

$$\hat{y}_i = F_M(x_i) = \sum_{m=1}^{M} h_m(x_i) \tag{3.25}$$

Where the h_m are fixed-size decision trees, the constant M is the number of estimators. The trees are built greedily:

$$F_m(x) = F_{m-1}(x) + h_m(x), \tag{3.26}$$

where the loss function defines $l(y_i, F(x_i))$. By applying a first-order Taylor approximation, the value l can be approximated as follows:

$$l(y_i, F_{m-1}(x_i) + h_m(x_i)) \approx l(y_i, F_{m-1}(x_i)) + h_m(x_i) \left[\frac{\partial l(y_i, F(x_i))}{\partial F(x_i)} \right]_{F=F_{m-1}}. \tag{3.27}$$

$\left[\frac{\partial l(y_i, F(x_i))}{\partial F(x_i)}\right]_{F=F_{m-1}}$ is the derivative of the loss function evaluated at $F_{m-1}(x)$ allowing it to be less computationally demanding. The derivative is denoted as g_i and by removing the constant terms, it yeilds:

$$h_m \approx \arg\min_h \sum_{i=1}^{n} h(x_i) g_i \tag{3.28}$$

In an iterative process, $h(x_i)$ when it predicts negative value negative gradient $-g_i$.

GB algorithms work well for heterogeneous, small, and non-linear datasets [60]. The library chosen for the GB analysis was *CatBoost*, developed Yandex by in 2017. *CatBoost* outperformed other boosting libraries in benchmarks, deals automatically with missing numerical values, handles categorical features, runs in GPU, and is relatively faster than other libraries [60].

3.8 Conclusion

The processes required to deal with data and develop the models were described in Chapter 3. The process of making a regression analysis was explained, along with the tools utilized in this research. After the models were built, the results could be computed and analyzed. The following chapter aims to provide an overview of the main findings from the data preprocessing. Later, the results from performance and sensitivity analysis are shown.

4

Results

The result chapter is divided into the three main stages of the methodology, shown in Figure 1.1. The stages were expanded to provide a better description of the results found, as shown in Figure 4.1.

Stage 1 describes the main findings from the data preprocessing and feature engineering in Section 3.1. The results from the correlation analysis and the final list of parameters are provided, followed by an overview of the data used to implement the models.

In Stage 2, several steps were defined to determine the most suitable model for the sensitivity analysis. With the initial performance, the type of analysis was selected to proceed to the optimization, and the best-performing tool was compared to the statistical model based on its accuracy.

In Stage 3, the SHAP values for the machine learning model were calculated, and the influence of the different parameters was evaluated, as described in Section 3.6. The sensitivity analysis considered the general impacts of the features for the models, the hypotheses collected from NL-LAB, highlighted in Section 2.2, and the differences between field and laboratory results.

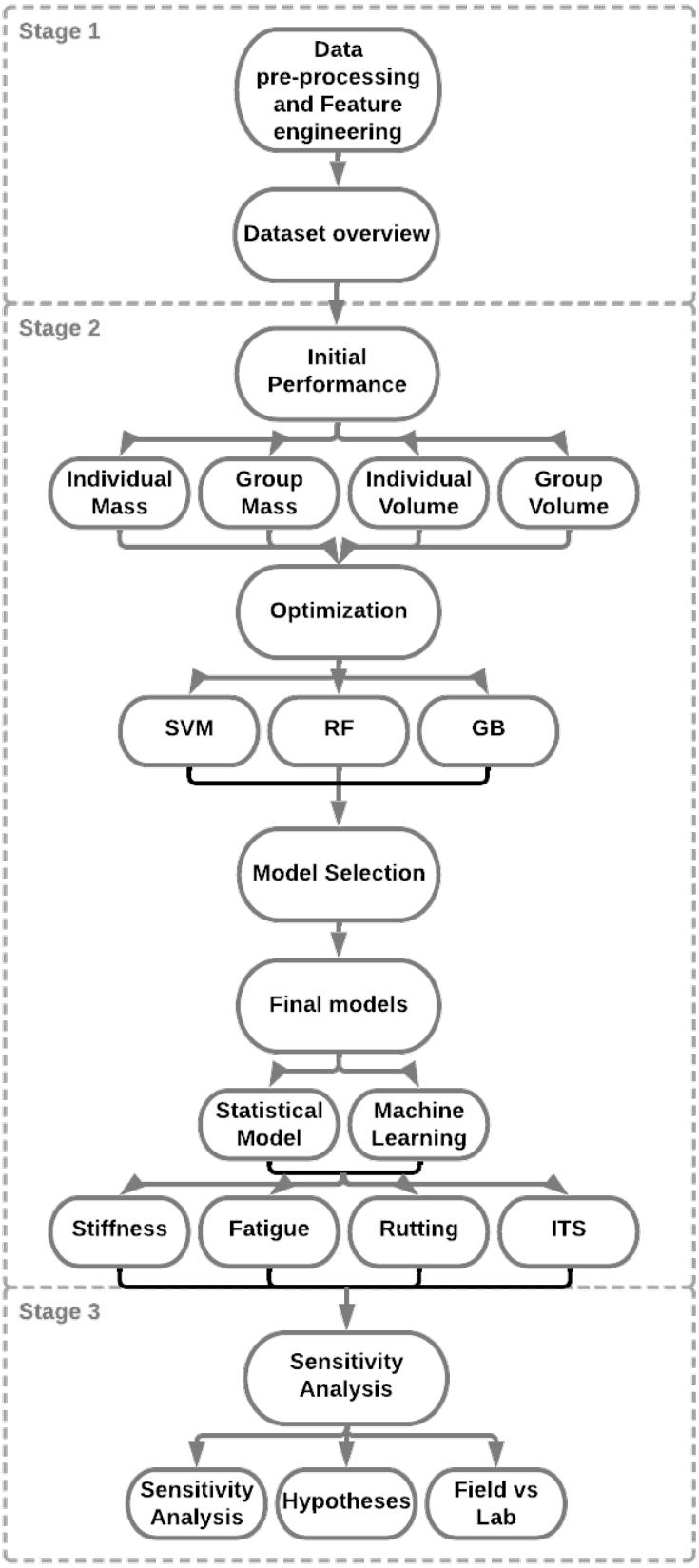


Figure 4.1: Results structure

4.1 Data preprocessing and feature engineering

With the data preprocessing and feature engineering, the final dataset was obtained. Table B.1 shows the entire extracted dataset. The filtering process removed all the features in Table B.2. From the removed features, bitumen properties and RAP percentage data were incorporated with the properties of another parameter. RAP percentage considered two different sources and the sum of the sources. Therefore, in the model the sum of the sources were considered. Additionally, tests were conducted to determine the bitumen properties of the fresh binder and a combination of the fresh and reclaimed binder. The properties found for the compound binder incorporated the properties of the fresh binder. Thus, the model considered only the properties of the combined mixtures.

As a final step, the phases and works parameters were used to understand the differences between the laboratory and the field. The works were the different roads on which the samples originated.

The preprocessing and feature engineering results are presented as follows.

4.1.1 Correlation Analysis

The purpose of the correlation analysis, presented in Figures B.1 and B.2, was to allow the removal of four parameters because of correlations above 0.8 as described in Section 3.1. The selection was the following:

Bitumen correlation:

- Ring and Ball was highly correlated to all bitumen properties. Thus it was removed.
- Complex modulus (G*) was highly correlated to three properties, and phase angle (δ) was highly correlated to two properties. For this reason, the G* was removed from the given data.

Density correlation:

- Density was highly correlated to air void content, and compaction degree also incorporated the density. Thus the density was also removed.
- Air void content was highly correlated to density, and compaction degree, and 74 missing values were missing. Thus the air voids was removed in the final analysis.

4.1.2 Final selection of parameters

The final list of parameters in the analysis were divided into dependent and independent variables. The dependent variables are predicted based on the independent variables. In this

research, the dependent variables were stiffness, ε_6 individual, f_c lin, and ITS. The final selection of independent variables is described in Table 4.1.

Table 4.1: Model's Variables

Dependent variables	
Name	**Description**
Stiffness	Stiffness @20°C and 8 Hz [MPa]
$\varepsilon_6(Individual)$	Individual calculated ε_6 value
f_c lin	Creep rate $[\frac{\mu\varepsilon}{cycles}]$
ITS	Indirect tensile strength @ 15°C [MPa]
Independent variables	
Name	**Unit**
Year	–
Bitumen penetration	$10^{-1}mm$
Compaction degree	%
Target density	$\frac{kg}{m^3}$
Phase angle	°
RAP percentage	%
Sample condition	–
Mixing setup	–
Compaction setup	–
Friction reduction system	–
Target filler percentage	%
Target bitumen percentage	%
Target percentage of aggregates retained in 22.4mm sieve	%
Target percentage of aggregates retained in 16mm sieve	%
Target percentage of aggregates retained in 11.2mm sieve	%
Target percentage of aggregates retained in 8mm sieve	%
Target percentage of aggregates retained in 5.6mm sieve	%
Target percentage of aggregates retained in 2mm sieve	%
Target percentage of aggregates retained in 0.063mm sieve	%

It is important to highlight that the aggregates and binder percentages were mass percentages. The assumed densities of binder and aggregates allowed to transform the mass percentages into volume percentages which were used to develop four different models. The target aggregate percentage was divided into individual and group components to understand their impact on the model's performance. All individual fractions of aggregates were considered in the individual components, and target stone and target sand percentages were considered for group components. The target stone percentage was determined by the sum of aggregates retained from the sieve 22.4mm to 2mm, and the target sand percentage was considered as the aggregates retained in the 63mμ sieve. The following acronyms describe these percentages:

- IM: Individual Mass
- IV: Individual Volume

- GM: Group Mass
- GV: Group Volume

It is highlighted that the analysis of the individual components consisted of all available independent features, excluding sample condition and friction reduction system. Sample condition was used exclusively for ITS and friction reduction system for permanent deformation. The total number of independent features for stiffness and fatigue was 17 for the individual component analysis and 12 for the group analysis. Permanent deformation and ITS were composed of 18 independent features for the individual component analysis and 13 for the group.

The final sensitivity analysis was incorporated with the Phases. The aim was to understand the differences between the field and the laboratory. The mixing setup, compaction setup, and year were removed from the last analysis because their values were incorporated into the Phases.

4.2 Data overview

The number of available data points for each functional property was different. Table 4.2 shows the available data points, and Table 4.3 gives an overview of the data.

Table 4.2: Total number of data points

Datapoints	**Stiffness**	ε_6 **Individual**	f_c **lin**	**ITS**
Total	425	407	120	192
Training set	340	326	96	154
Testing set	85	810	24	38

Table 4.3: Data overview

Variable	**Minimum value**	**Maximum value**
Year	0	6
Target density	2360	2399
Compaction degree	97.35	103.31
RAP Percentage	50	65
Bitumen Penetration	11	53
Bitumen Phase Angle	39.96	66.57
Target Mass composition Stone	52.58	57.9
Target Mass composition Sand	32.12	36.81
Target Mass composition Filler	5.72	7.64
Target Mass composition Bitumen	4.25	5.4
Mixing set-up	n/a	n/a
Compaction set-up	n/a	n/a
Friction Reduction System	n/a	n/a
Dry / Wet	Dry (0)	Wet (1)

The categorical data was composed of multiple unique values, especially mixing setup, compaction setup, and friction reduction system. Two values were assigned for sample condition as presented in the last row of Table 4.3. Dry samples were defined as zero and wet samples were defined

as one. Table 4.4 presents a data overview of the mixing and compaction setup, represented by columns and rows.

Table 4.4: Mixing and compaction setup data points

Set-up	Asphalt Plant	Forced Action Mixer	Planetary Mixer
Field roller	361	-	-
Gyrator comp.	69	43	35
Hand roller	6	12	-
Mini roller	27	18	9
Segment comp.	76	62	-
Shear box	10	9	-

The compaction and mixing setups were also represented by the Phases. All the data points related to the forced action mixer and planetary mixer were considered in Phase 1. All the data points from the asphalt plant, excluding the field roller, were part of Phase 2. The field roller and asphalt plant were included Phase 3. Table 4.5 presents the total data points used for each Phase.

Table 4.5: Phases data points

Phase 1	Phase 2	Phase 3
188	188	361

Table 4.6 presents the different friction reduction systems used to determine the creep rate.

Table 4.6: Friction reduction system

Friction reduction system	Data points
Two-layer rubber with silicon grease	79
2x Marshallpaper	17
PTFE gecoat vlies	12
Acre system	12

The following section presents the results that were obtained from the initial analysis.

4.3 Initial performance evaluation

The process of modeling a machine learning model can be time-consuming, mainly because of the optimization stage [70]. Hence, the models were trained with the default hyperparameters to avoid a long computational optimization process and to select one of the four analyses described in Section 4.1.2. The statistical analyses were performed following the same steps. It is important to highlight that the results shown in Table 4.7 were also the final R^2 from the statistical analysis.

Table 4.7: R^2 of the statistical models

	Stiffness	ε_6 **Individual**	$\mathbf{f_c}$ **lin**	**ITS**
IM	0.92	0.81	0.88	0.82
IV	0.92	0.81	0.88	0.82
GM	0.92	0.79	0.89	0.84
GV	0.92	0.79	0.9	0.83

The machine learning default hyperparameters provided R^2 values superior to 0.70 for most of the models, which correspond to a good accuracy. Although, some models showed poor performance as well. For instance, RF performed poorly for the fatigue analysis, and SVR did not perform well for the rutting analysis. Stiffness models provided the highest R^2, with minimum values of 0.94 and a maximum of 0.95, as shown in Table 4.8.

Table 4.8: R^2 of the machine learning models

	Stiffness			ε_6 **Individual**			$\mathbf{f_c}$ **lin**			**ITS**		
	SVM	RF	GB	SVM	RF	GB	SVM	RF	GB	SVM	RF	GB
IM	0.94	0.94	0.94	0.8	0.63	0.71	0.3	0.8	0.75	0.79	0.72	0.73
IV	0.94	0.94	0.94	0.81	0.64	0.71	0.33	0.8	0.81	0.83	0.74	0.74
GM	0.94	0.95	0.94	0.82	0.64	0.72	0.39	0.81	0.77	0.8	0.75	0.75
GV	0.94	0.95	0.94	0.82	0.63	0.7	0.41	0.8	0.82	0.79	0.74	0.71

Tables 4.7 and 4.8 show that the performance of different models are similar. The highest variation in accuracy between the group and individual components was below 10%. Therefore, GM was selected to proceed with the optimization process.

4.4 Optimization and cross-validation

The optimization process was the same for all the different tools used. However, different tools use different sets of hyperparameters. Only the parameters listed below were chosen to reduce the computational demands of the analysis. In the optimization process, the models were generalized using 10-fold cross-validation. The computed R^2 represented the model's accuracy, evaluating the training set, cross-validation of the training set and the testing set. It is noted that the cross-validation accuracy was considered as the average of each fold.

The following subsections present the hyperparameters that were optimized.

4.4.1 Support Vector Regression

The SVR model was developed with the following four hyperparameters:

- Kernel: Defines if a linear or non-linear model could be applied.
- C: Penalty parameter of the error.
- γ: σ from the RBF kernel. Define the curvature of the function.

- ε: Size of the ε-Insensitive tube.

The application of the RBF kernel resulted in the best performance for SVR in the given dataset, which means that the non-linear model outperformed the linear. Most of the accuracies, presented in Table 4.9, were higher than 70%, which was satisfactory. Although, the SVR model poorly fitted the creep rate, poorly performing for the training and testing set. However, the high accuracy found in the cross-validation shows that the model could fit different distributions. Table 4.9 presents the resulting hyperparameters and R^2.

Table 4.9: SVR hyperparameters and R^2

	Stiffness	ε_6 **Individual**	$\mathbf{f_c}$ **lin**	**ITS**
Kernel	RBF	RBF	RBF	RBF
C	1.5	1.5	1.5	10
Gamma	0.1	0.2	0.3	0.1
ε	0.1	0.1	0.2	0.1
R^2_{train}	0.959	0.762	0.404	0.869
R^2_{test}	0.938	0.800	0.216	0.816
R^2_{CV}	0.896	0.756	0.732	0.796

4.4.2 Random Forests

Random forests have different hyperparameters when compared to SVMs. The list of parameters selected for tuning was as follows:

- Max depth: Maximum depth of the tree.
- Max Features: The number of features to consider when looking for the best split.
- Min samples leaf: The minimum number of samples required to be at a leaf node.
- Min samples split: The minimum number of samples required to split an internal node.
- Number of estimators: The number of trees in the forest.

Similarly to SVR, RF resulted in prediction accuracies higher than 70%, excluding the resistance to permanent deformation, as shown in Table 4.10. For the resistance to permanent deformation, the prediction for the training set and test set were higher than 70%. However, the cross-validation was significantly lower. The low quality of the prediction indicates that the model found patterns for the original split and did not find patterns for a new distribution. Table 4.10 shows the hyperparameters and R^2 from RF.

Table 4.10: RF hyperparameters and R^2

	Stiffness	ε_6 **Individual**	$\mathbf{f_c}$ **lin**	**ITS**
Max depth	80	80	80	80
Max features	5	2	5	5
Min samples leaf	3	3	3	3
Min samples split	8	8	8	8
Number of estimators	1000	1000	200	1000
R^2_{train}	0.954	0.723	0.815	0.867
R^2_{test}	0.943	0.756	0.729	0.743
R^2_{CV}	0.886	0.736	0.514	0.723

4.4.3 Gradient Boosting

GB model has a similar structure to RF, with similar hyperparameters. The list of parameters selected for tuning was as follows:

- Learning rate: Scales the contribution of each tree.
- Depth: Maximum depth of the tree.
- l2 leaf reg: Penalty parameter of the error.
- Min child samples: The minimum number of samples in a leaf.
- Odd type and odd wait: Over-fitting detectors.

The main difference between RF was the learning rate, which scales the predictions to advance toward the target value in a stepwise manner. It is important to highlight that for gradient boosting, over-fitting detectors were applied (od type and od wait) [90]. These over-fitting detectors were hyperparameters set to stop the analysis for a number of iterations that did not show improvement. Stopping prevents the model from "memorizing" the data because the model does not proceed with the training while no improvement is visible. "Od type" and "od wait" were set to "Iter" and 20, respectively. The "Iter" command stops training for a specified number of iterations, and "od wait" is the specified number of iterations.

The results found by the model were good and consistent with similar findings for the training, testing, and cross-validation, indicating that the model could capture the patterns for different datasets. In addition, as described at the beginning of this section, GB can add the trees into a single model, the models created in the cross-validation were added and the final prediction was the average of all models. Table 4.11 presents the hyperparameters and R^2 from GB.

Table 4.11: GB hyperparameters and R^2

	Stiffness	ε_6 **Individual**	$\mathbf{f}_c$ **lin**	**ITS**
Learning rate	0.016	0.01	0.02	0.009
Depth	10	10	13	11
l2 leaf reg	2.5	1	1	4
Min child samples	8	32	1	32
R^2_{train}	0.954	0.829	0.895	0.900
R^2_{test}	0.951	0.802	0.812	0.824
R^2_{CV}	0.919	0.762	0.794	0.823

The next step was the definition of the model to be further investigated. This selection is necessary to narrow the analysis and focus on sensitivity analysis.

4.5 Model Selection

Final models were selected on the basis of accuracy and interpretability. However, more aspects could be considered, for instance, time and preprocessing demands. Although the

time and preprocessing were out of the scope of this research, it is important to highlight the differences between the models. SVR training time was the fastest in all models with the highest preprocessing demands. All the numerical features needed to be scaled and transformed, and the categorical data encoded. GB required the least preprocessing and a longer optimization process.

Comparing the R^2 values in Tables 4.9, 4.10 and 4.11, GB outperformed all the models in most of the analyses. In general, it was able to capture patterns of the dataset and make a better prediction. Moreover, comparing the different sets of data, GB provided balanced results. Thus, GB was selected to proceed to the sensitivity analysis. Table 4.12 provides a summary of the best-performing models for the training, testing, and cross-validation processes.

Table 4.12: Performance comparison

	Stiffness	**ε_6 Individual**	**f_c lin**	**ITS**
R^2_{train}	SVM	GB	GB	GB
R^2_{test}	GB	GB	GB	GB
R^2_{CV}	GB	GB	GB	GB

The accuracy was computed for different loss functions to check its impact. It is important to highlight that the R^2 presented in Table 4.13 resulted from the optimization process using RMSE and cross-validation with the different loss functions to avoid the long optimization process for all the different loss functions. Stiffness and resistance to fatigue models performed the best with RMSE. ITS performed the best using quantile loss and rutting with expectile loss. The accuracy increased by approximately 1% and 0.1%, respectively. The change in performance for the different loss functions was negligible. Table 4.13 presents the accuracy for the different loss functions:

Table 4.13: Performance for different loss functions

Loss function	**Stiffness**	**ε_6 Individual**	**f_c lin**	**ITS**
RMSE	0.951	0.802	0.812	0.824
Quantile	0.951	0.798	0.796	0.835
Expectile	0.949	0.79	0.813	0.814

In summary, the selected models used gradient boosting with GM as inputs. The following section discusses the statistical and machine learning models' prediction plots and highlights the importance of hyperparameters.

4.6 Selected model

This section highlights the performance of the statistical and machine learning models, followed by the importance of hyperparameters. Next, the sensitivity analysis is presented.

4.6.1 Stiffness model

4.6.1.1 Predictive performance

The statistical model and machine learning model showed very high prediction accuracy above 90% ($R^2_{Test,GB} = 0.95$ and $R^2_{Test,MLR} = 0.91$), presented in Figures 4.2 and 4.3. GB's predictive accuracy was superior to MLR, better capturing the data patterns. In GB, the "min child samples" was found to be the most relevant hyperparameter, followed by the tree depth, learning rate, and l2 leaf reg, as shown in Figure 4.4.

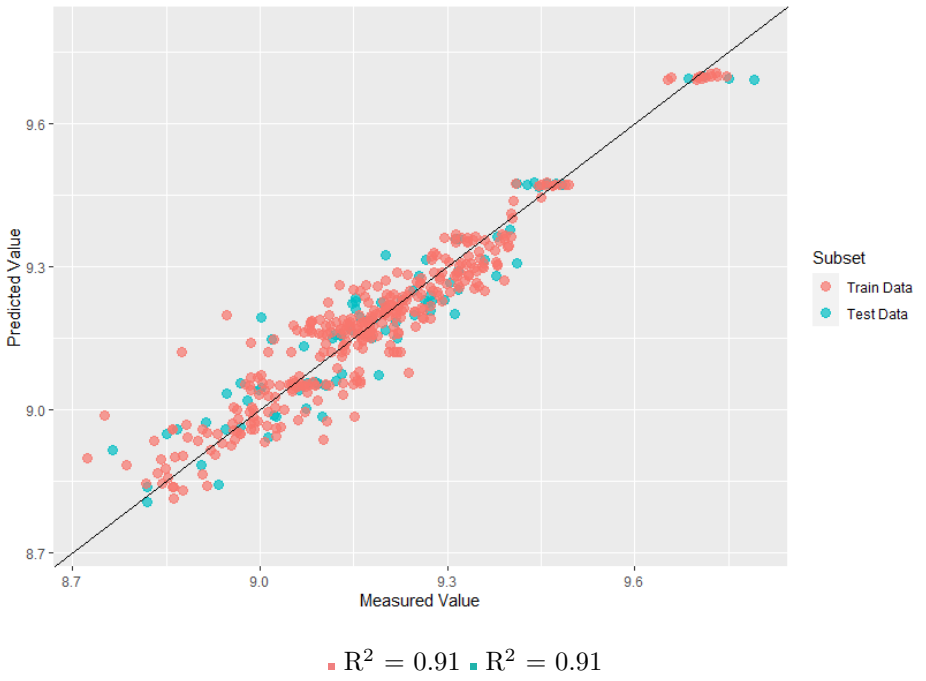


Figure 4.2: Predicted-measured stiffness for MLR predictive model.

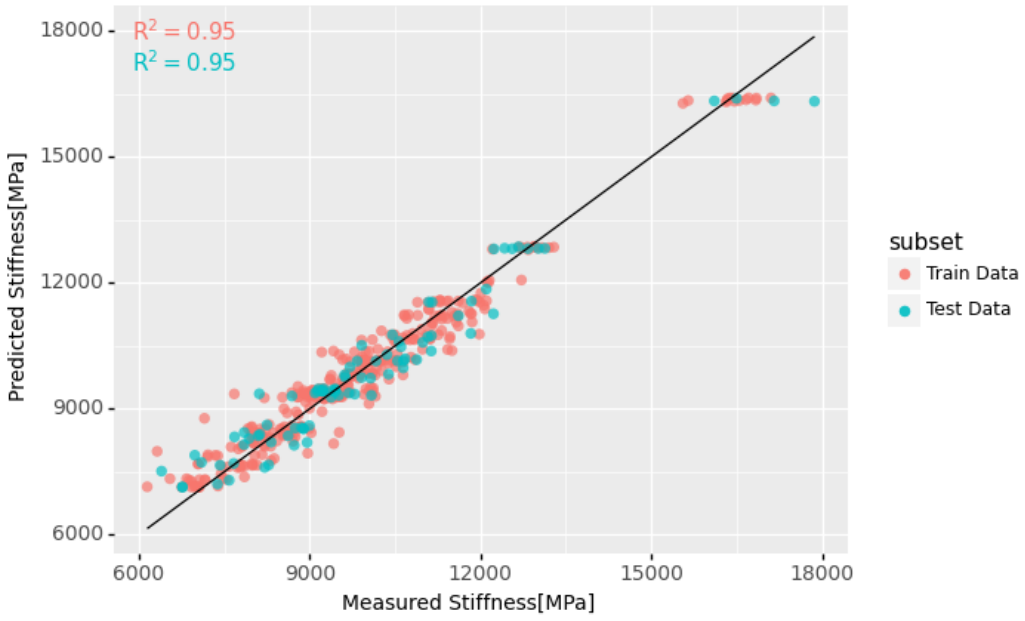


Figure 4.3: Predicted-measured stiffness for GB predictive model.

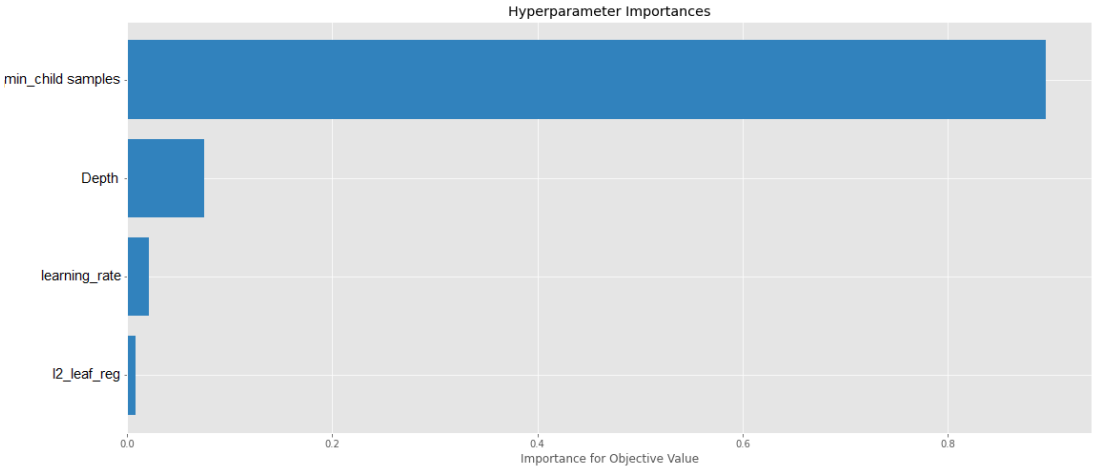


Figure 4.4: Stiffness hyperparameter importance

4.6.1.2 General sensitivity analysis

The most important parameters for stiffness were found to be compaction degree, phase angle, and mixing setup. Figures 4.5, 4.6, and 4.7 show the feature importance and their impact on the model.

Compaction degree was found to be the most important parameter in the analysis, as shown in Figure 4.5. It relates to the density and air void content, as a higher density generally results in higher stiffness [23].

The phase angle is part of the complex modulus, which is expected to have high importance to the stiffness because of its high correlation to stiffness, as described in Chapter 2 [16]. Additionally, materials with a higher phase angle can deform more easily and, therefore, be less stiff.

In the mixing setup, the planetary mixer only positively influenced the stiffness, as shown in Figure 4.6. However, the number of available data points was small to formulate a concrete conclusion. The forced action mixer had similar data points impacting the stiffness positively and negatively. Asphalt plants also had positive and negative impacts on stiffness. However, the asphalt plant presented the most reliable conclusion because it corresponded to the most available data points.

The results for some features disagreed with the findings in the literature review or were inconclusive. Increasing the RAP content reduced the stiffness when it was expected to increase [45]. The results for the target composition were found to be inconclusive. The low variability of the data may not allow the model to capture their effects.

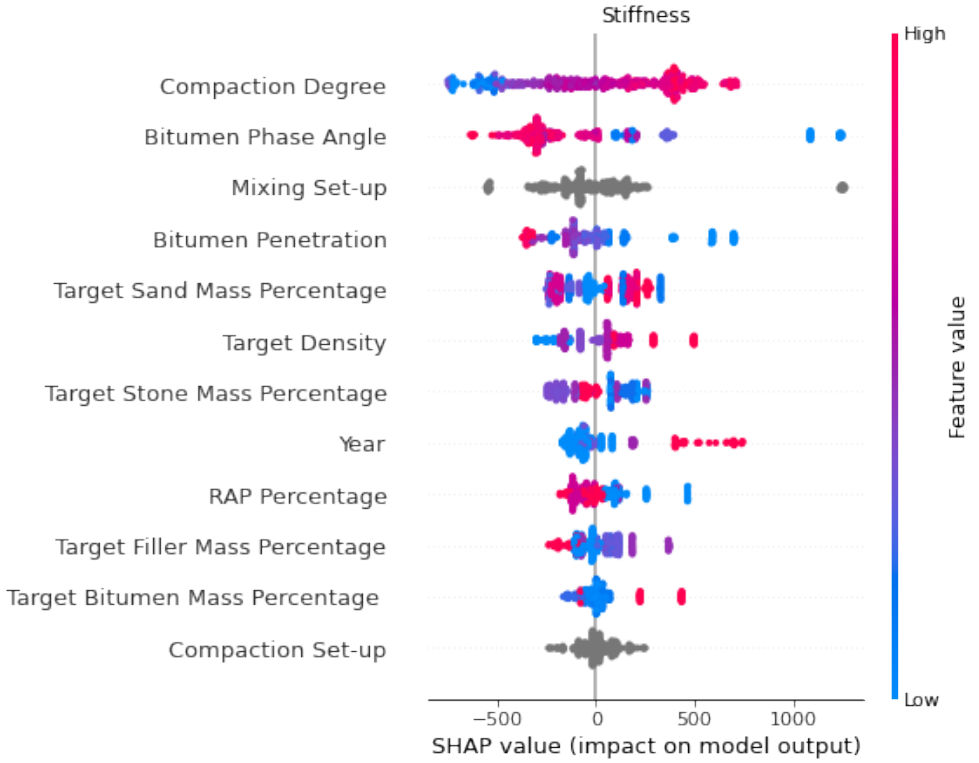


Figure 4.5: Summary SHAP plot

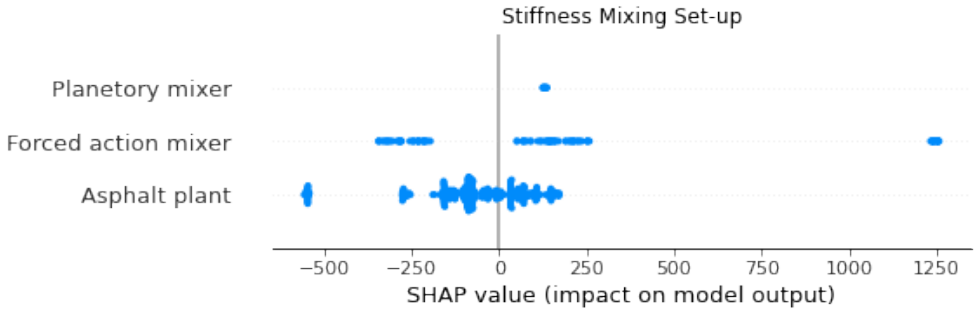


Figure 4.6: Summary SHAP plot for mixing setup

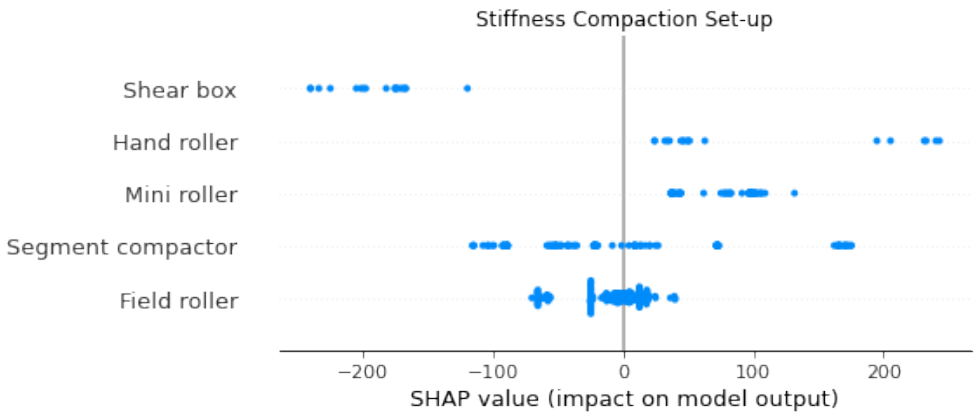


Figure 4.7: Summary SHAP plot for compaction setup

4.6.1.3 Hypotheses

The stiffness was divided into three main hypotheses determined in the NL-LAB, presented in Section 2.2. Table 4.14 summarizes the hypotheses:

Table 4.14: Stiffness Hypotheses

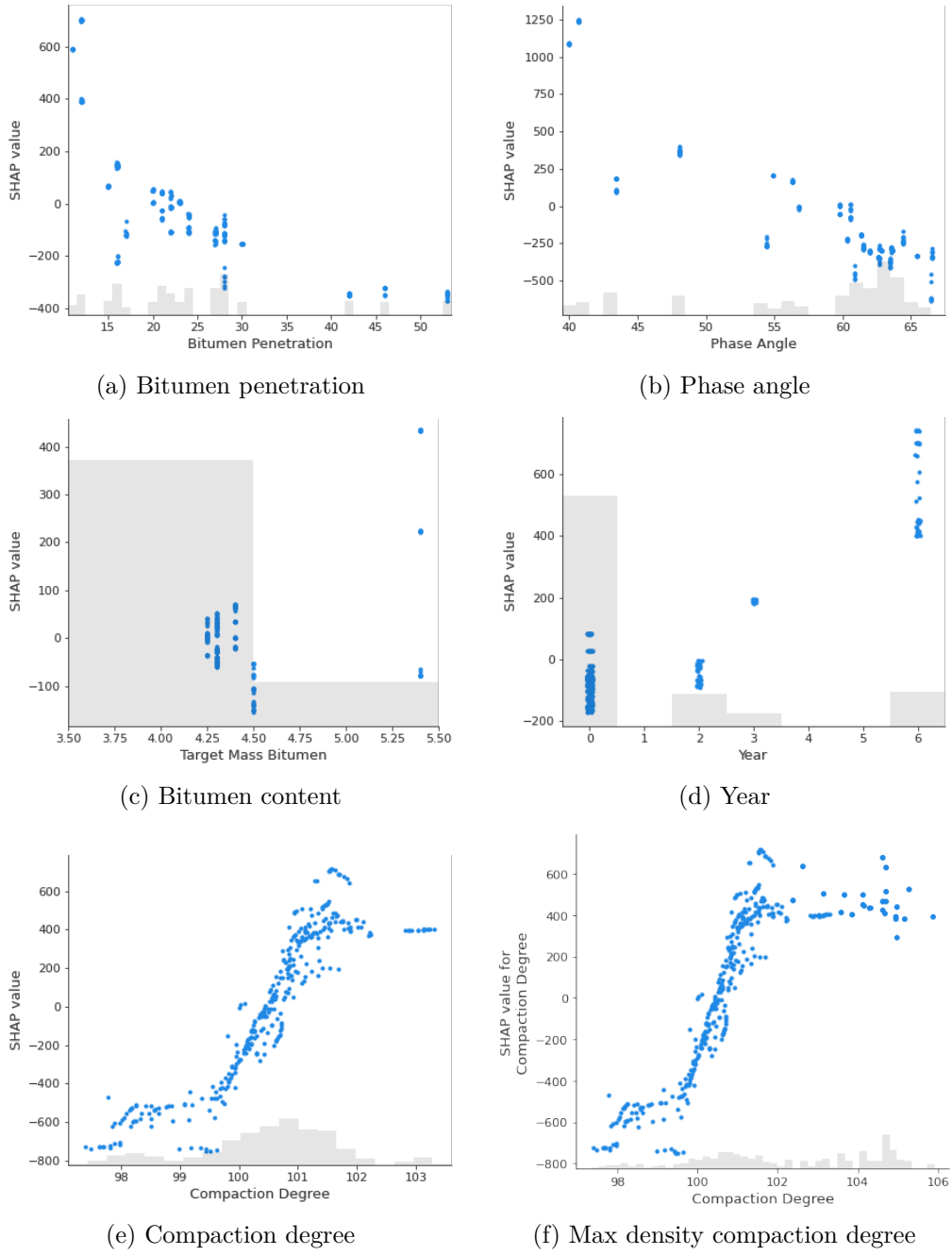
Hypothesis	Effect on the property
More and softer bitumen	Lower/Smaller
Higher density	Higher/Larger
Aging	Higher/Larger

Figure 4.8 a to c shows the bitumen content and its tested properties. Figure 4.8 a and b shows that a higher penetration and phase angle results in a lower stiffness. It is noted that the stiffness of the binder is related to the overall stiffness of the mixture [37]. The phase angles crossing the 45° margin reach the flow point, and the binder behaves more fluidly, which reduces the stiffness once the binder stops behaving solid-like and moves towards a fluid behavior. Figure 4.8 c shows that a higher percentage of bitumen leads to higher stiffness. However, the data points for the highest bitumen content showed opposite results. The bitumen content data with 5.4% belongs to Work 6, which was polymer modified, which could be the reason for the difference. Lower bitumen content results in stiffer mixtures. It is important to highlight that bitumen content had low importance for the model.

Figure 4.8 d shows that the stiffness increases over the years. The aging of bituminous materials leads to an increase in their stiffness [42].

Figure 4.8 e and f are plots based on the compaction degree and compaction degree considering maximum density. Higher density results in higher stiffness, which was expected [23]. However, the positive impact has limitations. For 101.5% compaction degree, the contribution to stiffness reached a maximum, and above, there was no improvement. By plotting the impact of the compaction degree with maximum density, the trend continued the same, meaning that there is a maximum possible contribution from density to the stiffness [24]. Moreover, the SHAP plot had an apparent sigmoidal shape, indicating that after a limit, the stiffness would not increase anymore and would maintain constant even with the increase in density.

From a data perspective, the model has shown that the hypotheses relating to stiffness were correct. Increasing the bitumen content and using softer binders reduces the stiffness. Increasing the density increases the stiffness, and aging results in higher stiffness. However, the sigmoidal trend suggests a limit to the positive influence of density on stiffness. The shape found for the compaction degree plot could be potentially used to determine the target density of mixtures. Fitting a sigmoidal function would make it possible to determine the point at which the density stops improving the stiffness.



(a) Bitumen penetration

(b) Phase angle

(c) Bitumen content

(d) Year

(e) Compaction degree

(f) Max density compaction degree

Figure 4.8: Individual SHAP values - Density

4.6.1.4 Phases analysis

The phase sensitivity analysis resulted in two central readings. First, the change in stiffness regarding the material's aging is evident. Figure 4.10 shows a clear trend from "Phase 3-Year 0", increasing the positive impact through the years. The second interpretation was the range of influence on stiffness for the different phases. The different Phases presented different impacts on the stiffness. However, the variability was small, meaning that the results found in the laboratory were similar to the field.

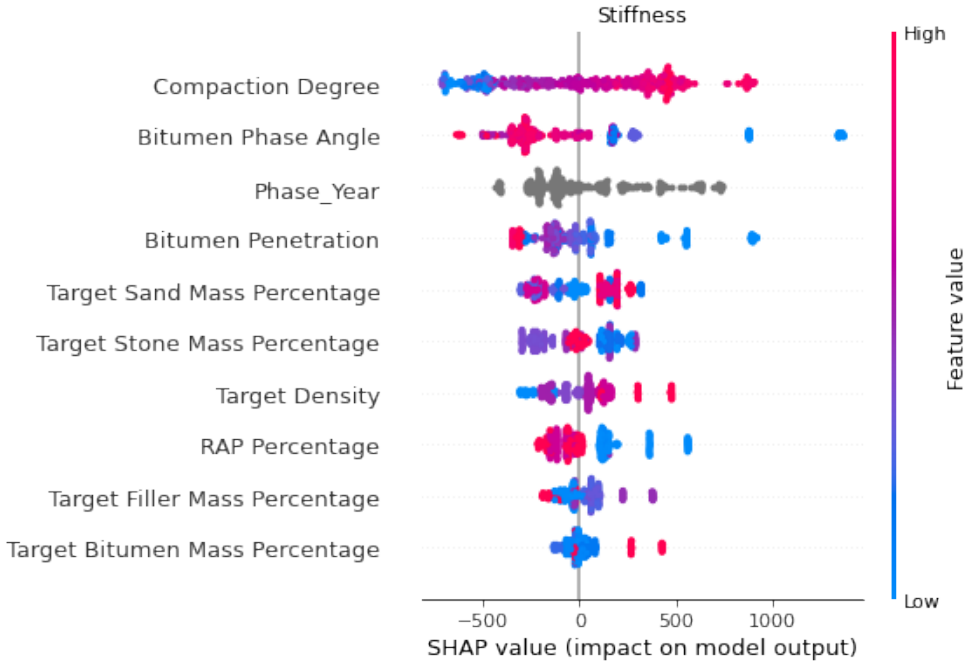


Figure 4.9: Summary SHAP plot

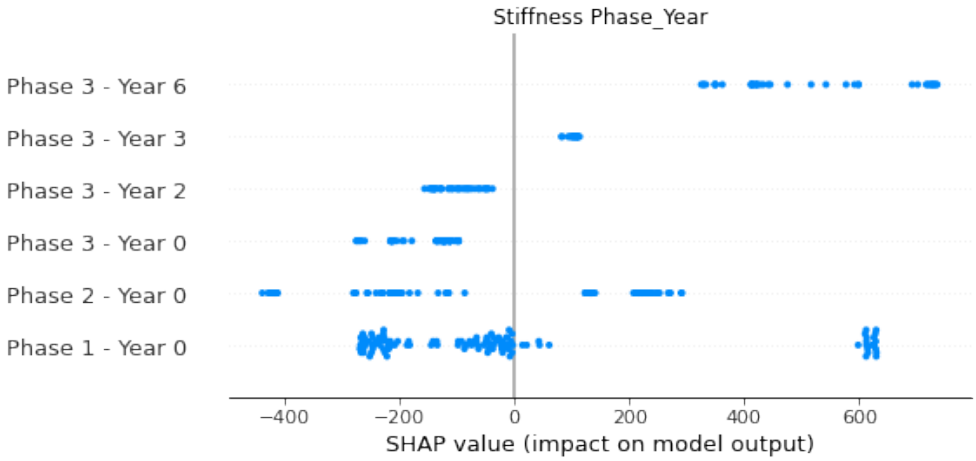


Figure 4.10: Summary SHAP plot for Phase

4.6.2 Resistance to fatigue

4.6.2.1 Predictive performance

The statistical and machine learning models showed good prediction accuracy. GB outperformed MLR by a 1% margin ($R^2_{Test,GB}$ = 0.81 and $R^2_{Test,MLR}$= 0.79), presented in Figures 4.11 and 4.12. Both MLR and GB provided the predictions spread in horizontal lines. The horizontal lines were a possible indication of the systematic errors caused by the tests and problems caused by the extrapolation. Fatigue lines are fit from the failure point of several samples, and this single line provides a general indicator for fatigue life. In reality, each sample would have different fatigue lives, and because of this generalization, the data points are spread in such a

manner. In GB, the "min child samples" was found to be the most relevant hyperparameter for the analysis. Followed by the tree depth, l2 leaf reg, and learning rate, as shown in Figure 4.13.

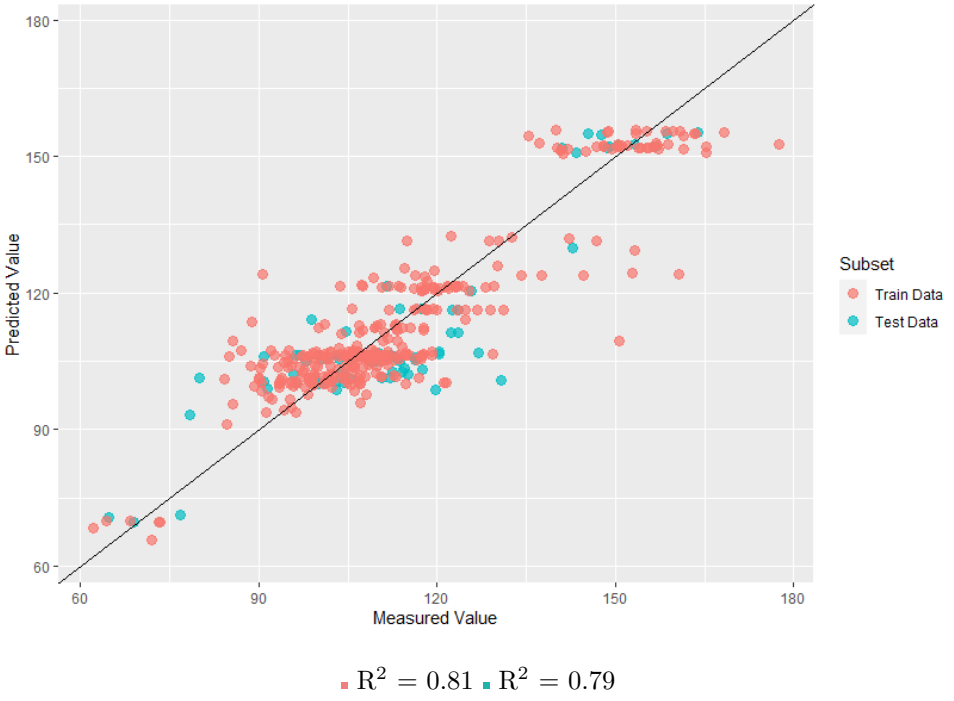


Figure 4.11: Predicted-measured ε_6 for MLR predictive model.

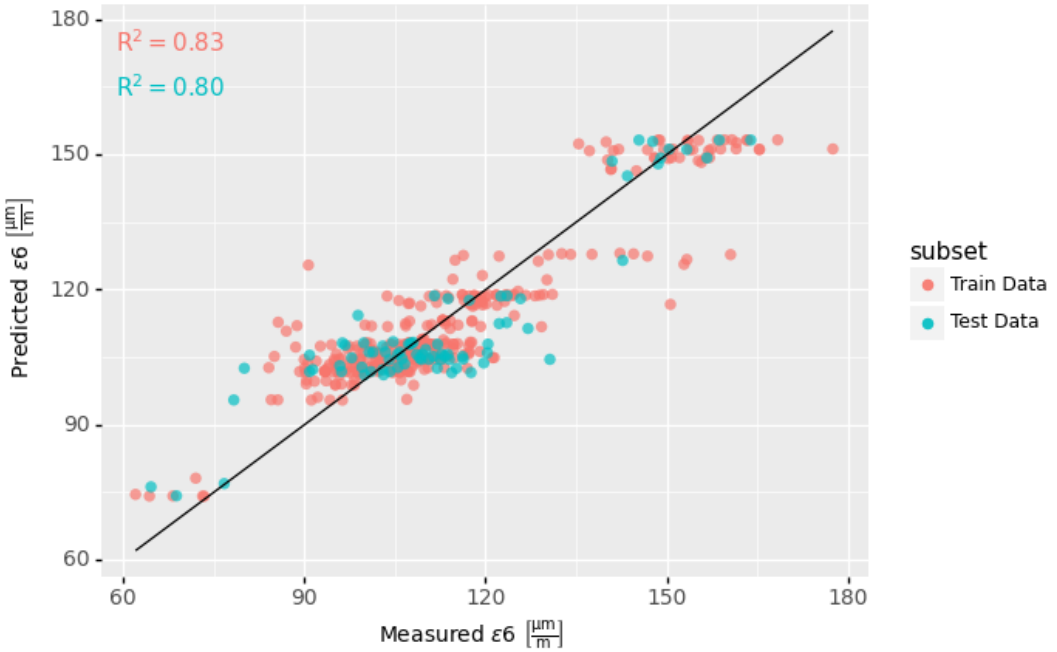


Figure 4.12: Predicted-measured ε_6 for GB predictive model.

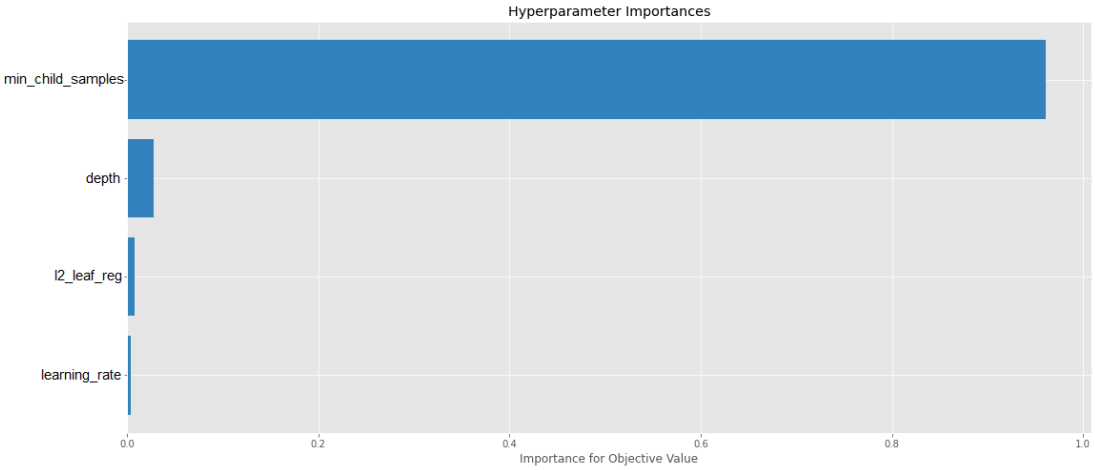


Figure 4.13: ε_6 hyperparameter importance

4.6.2.2 General sensitivity analysis

The main features affecting the fatigue resistance were the target bitumen mass percentage, target filler mass percentage, and target density. The combination of filler and bitumen results in the mastic [47]. The stiffness of the mastic is related to the fatigue life of the mixture. Higher mastic stiffness leads to higher fatigue life and reduces crack propagation [47].

The bitumen content influences the stiffness and brittleness of the mixture [4]. Mixtures with lower volume content are more brittle and more prone to crack. The results from Figure 4.14 show that a higher bitumen content results in higher resistance to fatigue.

The results shown in Figure 4.14 indicate that low percentages of filler content decrease fatigue resistance. However, the highest positive impact resulted from values below the maximum filler content. Higher filler content leads to a stiffer material. The excessive use of filler leads to brittleness compromising fatigue resistance [46].

The results from the target density indicate that higher values increase fatigue resistance. Higher densities affect the air void content, increasing the initial stiffness [24]. Higher stiffness suggests a higher fatigue life.

Similarly to stiffness, the target composition results were generally inconclusive. Target filler mass percentage had high importance to the model, indicating that even low variability allows capturing the patterns for features with high importance. The remaining target aggregate percentages were inconclusive.

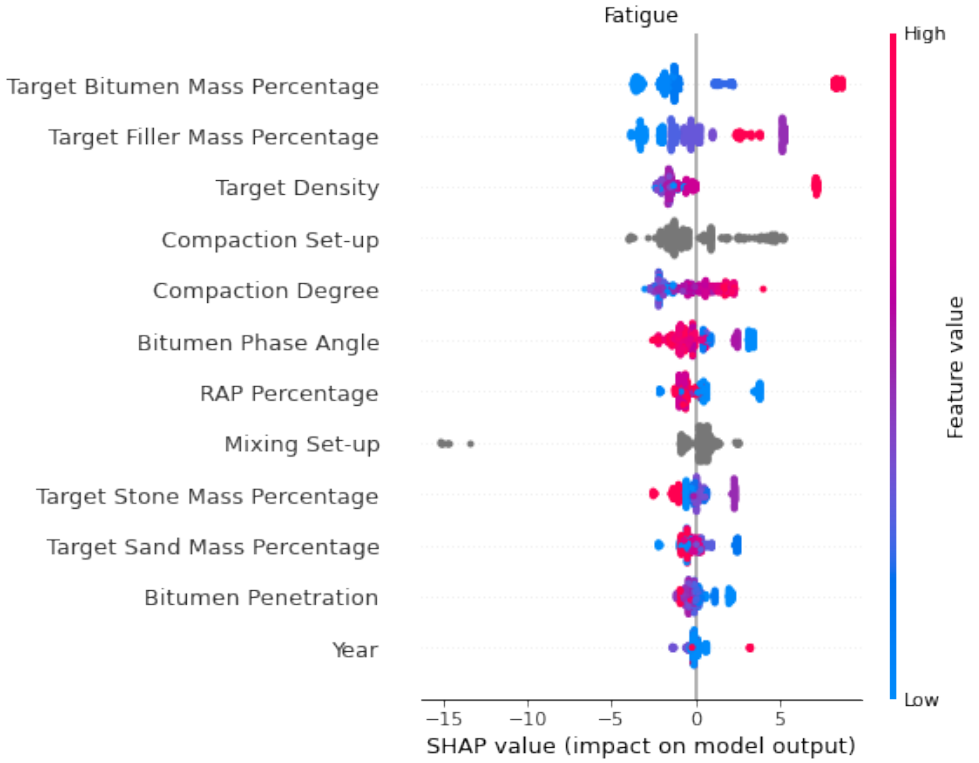


Figure 4.14: Summary SHAP plot

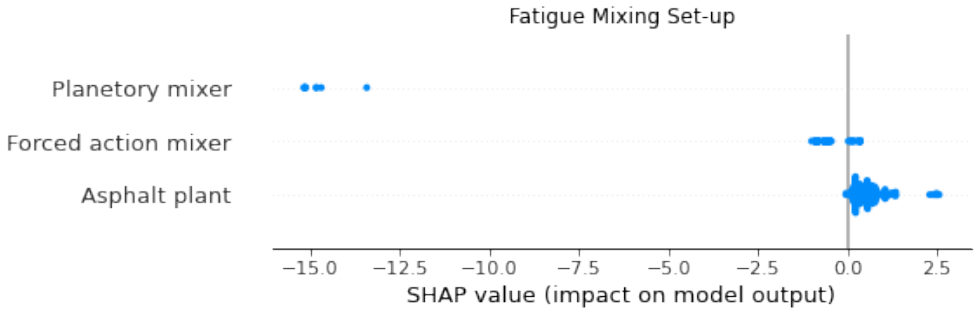


Figure 4.15: Summary SHAP plot for mixing setup

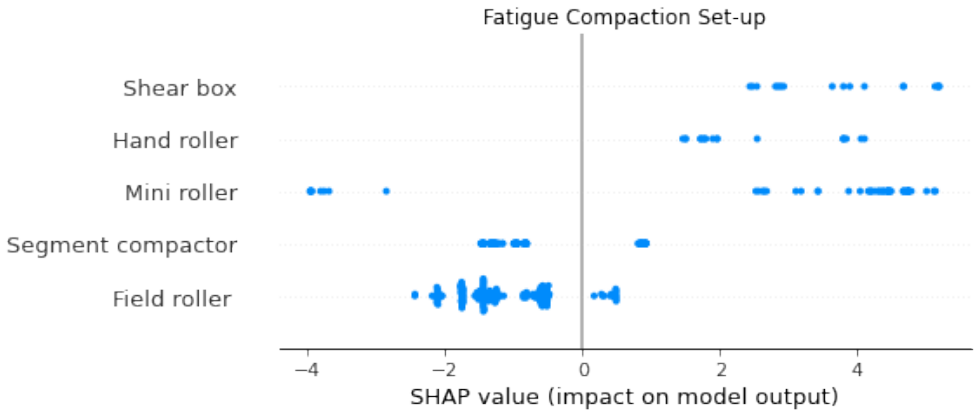


Figure 4.16: Summary SHAP plot for compaction setup

4.6.2.3 Hypotheses

The fatigue resistance was divided into three main hypotheses determined in the NL-LAB, presented in Section 2.2. Table 4.15 summarizes the hypotheses:

Table 4.15: Resistance to fatigue Hypotheses

Hypothesis	Effect on the property
More and softer bitumen	Higher/Larger
Higher density	Higher/Larger
Aging	Lower/Smaller

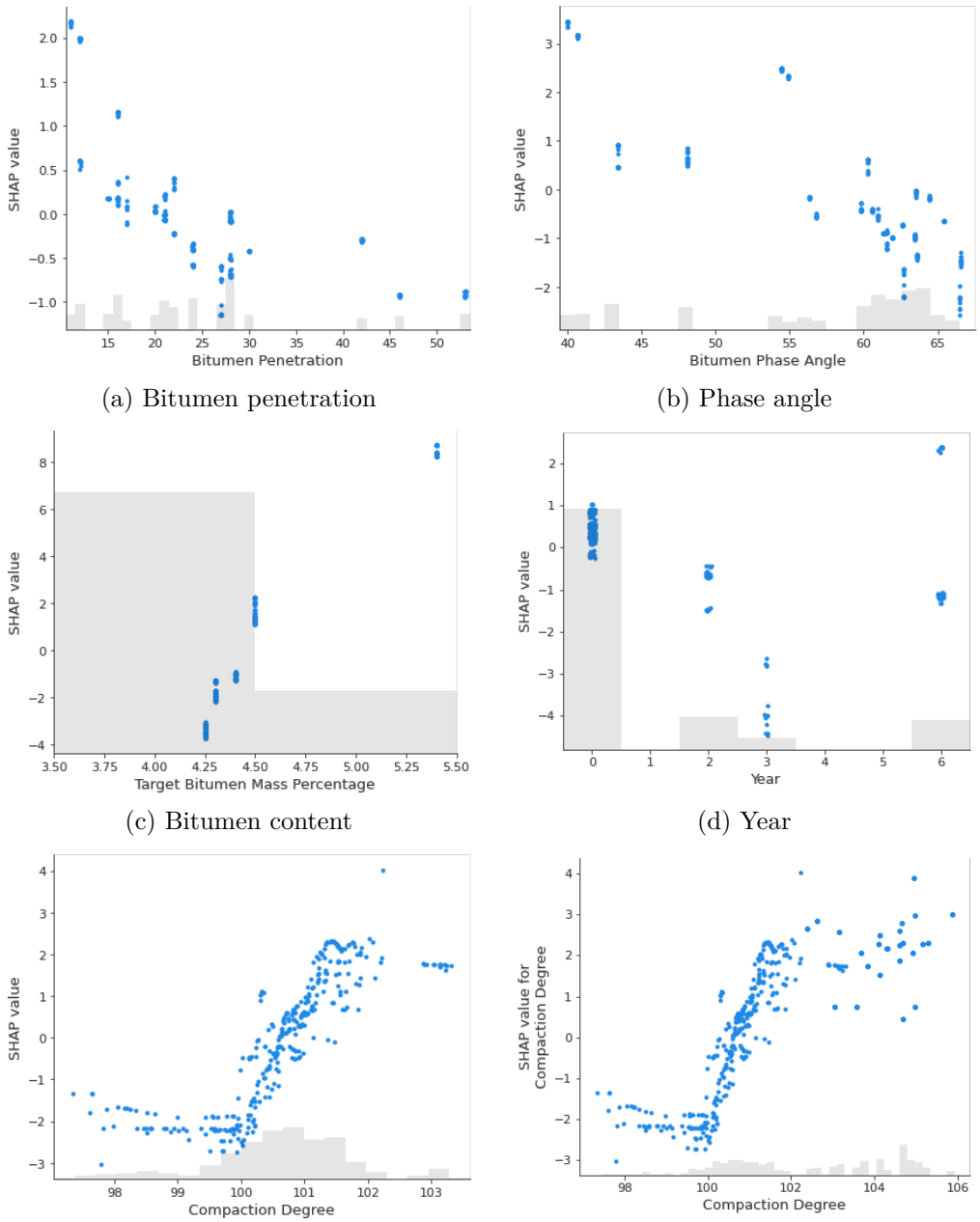
Figure 4.17 a to c shows the bitumen content and its tested properties. Figure 4.17 a and b show that higher penetration and phase angle results in lower resistance to fatigue. Softer binder reduces the stiffness of the mixture, also affecting the resistance against cracking [37]. Complementary, the increase of the phase angle reduces the elastic modulus of the mixture, resulting in a less stiff material reducing the resistance to fatigue. It is important to highlight that referring to Figure 4.14, bitumen penetration and phase angle were found to have a low overall contribution to the resistance to fatigue. Figure 4.17 c shows that higher bitumen content results in higher resistance to fatigue. Low bitumen content makes the mixture more prone to crack. Hence, higher content increase the resistance to fatigue [4].

Figure 4.17 d shows a downward trend with some outliers in years 3 and 6. The aging leads to a stiffer binder [37]. The increased stiffness of the aged binder leads to a higher cracking potential. Hence, lower resistance to fatigue [42]. It is important to highlight that referring to Figure 4.14, aging showed the lowest overall contribution to fatigue resistance.

Figures 4.17 e and f are plots based on the compaction degree and compaction degree considering maximum density. Higher density results in higher fatigue resistance. However, similar to stiffness, the positive impact has limitations. For 101.5% compaction degree, the contribution to the resistance to fatigue reached a maximum, and above, there was no improvement. The trend gets more scattered by plotting the impact of the compaction degree with maximum density but continued to develop asymptotically to the x-axis, meaning that, similarly to stiffness, there is a maximum possible contribution from density to fatigue resistance. Moreover, the SHAP plot showed an apparent sigmoidal shape, indicating that after a limit, the resistance would not increase anymore and would remain constant for increasing densities.

From a data perspective, the model showed results disagreeing with some hypotheses. Higher bitumen content improves fatigue resistance, while softer bitumen reduces it. The density and aging hypotheses were in agreement with the results found. Higher density leads to higher ε_6 and aging to a lower ε_6. Moreover, similarly to stiffness, the shape found for the compaction degree plot could be potentially used to determine the target density of mixtures. Fitting a

sigmoidal function would make it possible to determine the point at which the density stops improving the fatigue resistance.



(a) Bitumen penetration

(b) Phase angle

(c) Bitumen content

(d) Year

(e) Compaction degree

(f) Max density compaction degree

Figure 4.17: Individual SHAP values

4.6.2.4 Phases analysis

The change in fatigue regarding the aging of the material was evident. Figure 4.19 shows a clear trend from "Phase 3 - Year 0", increasing the negative impact in fatigue resistance through the years. Regarding the differences between Phases, the impact variability on ε_6 individual for Phase 1 and 3 is small, meaning that results from the laboratory did not diverge too much from the field. Phase 2 showed results with different ranges compared to Phases 1 and 3.

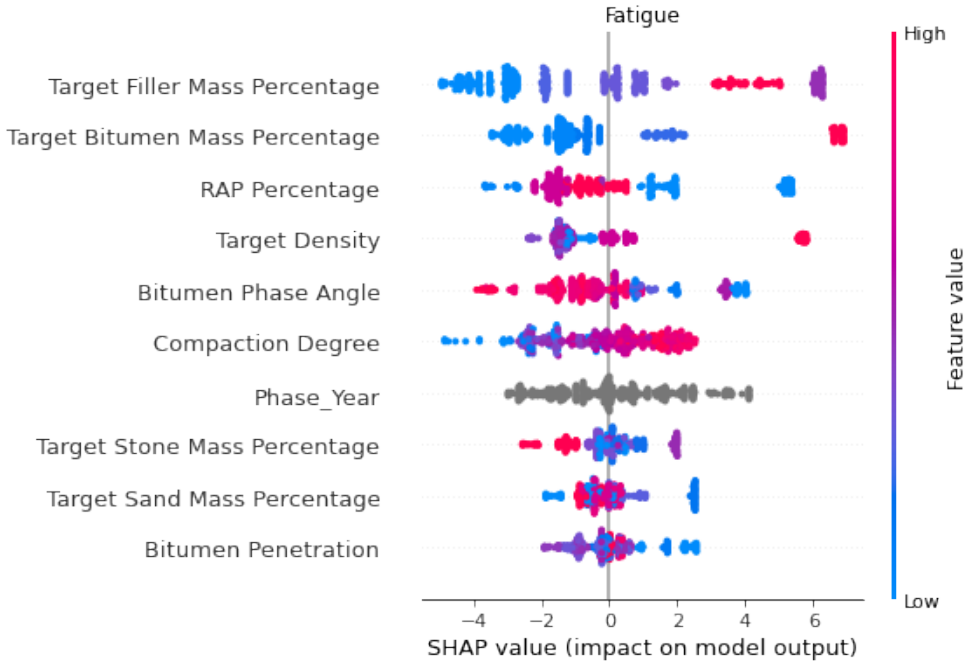


Figure 4.18: Summary SHAP plot

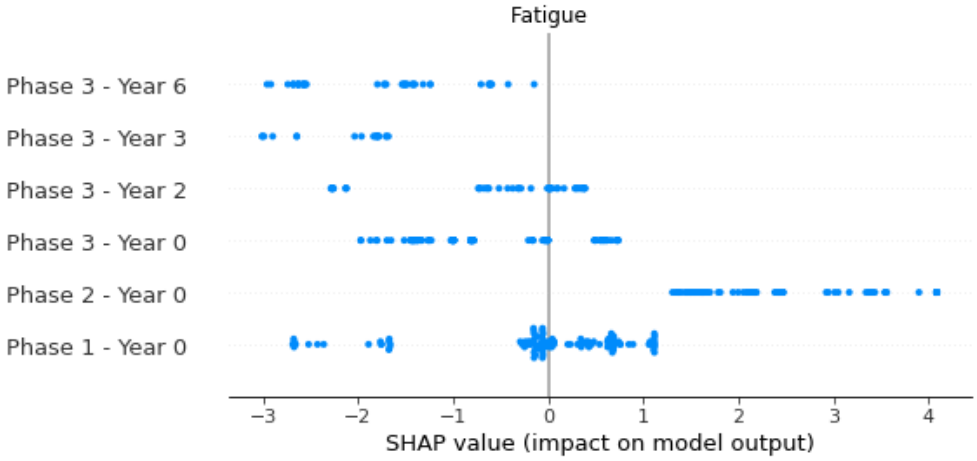


Figure 4.19: Summary SHAP plot for Phase

4.6.3 Resistance to permanent deformation

4.6.3.1 Predictive performance

The statistical and machine learning models showed good prediction accuracy for resistance to permanent deformation. MLR outperformed GB by 2% margin ($R^2_{Test,GB}$ = 0.81 and $R^2_{Test,MLR}$= 0.83), as shown in Figures 4.20 and 4.21. Figure 4.20 has a smoother distribution of the data points on the 45° line than machine learning. The machine learning model overestimated the predictions for low creep rate values, ranging from 0 to 0.12 $[\frac{\mu\varepsilon}{cycle}.10^6]$, from 0.25 to 0.37 the predictions were smooth, and for values above 0.35 the predictions were underestimated. In GB, the "min child samples" was found to be the most relevant hyperparameter for the analysis. Followed by the learning rate, l2 leaf reg, and depth, as shown in Figure 4.22.

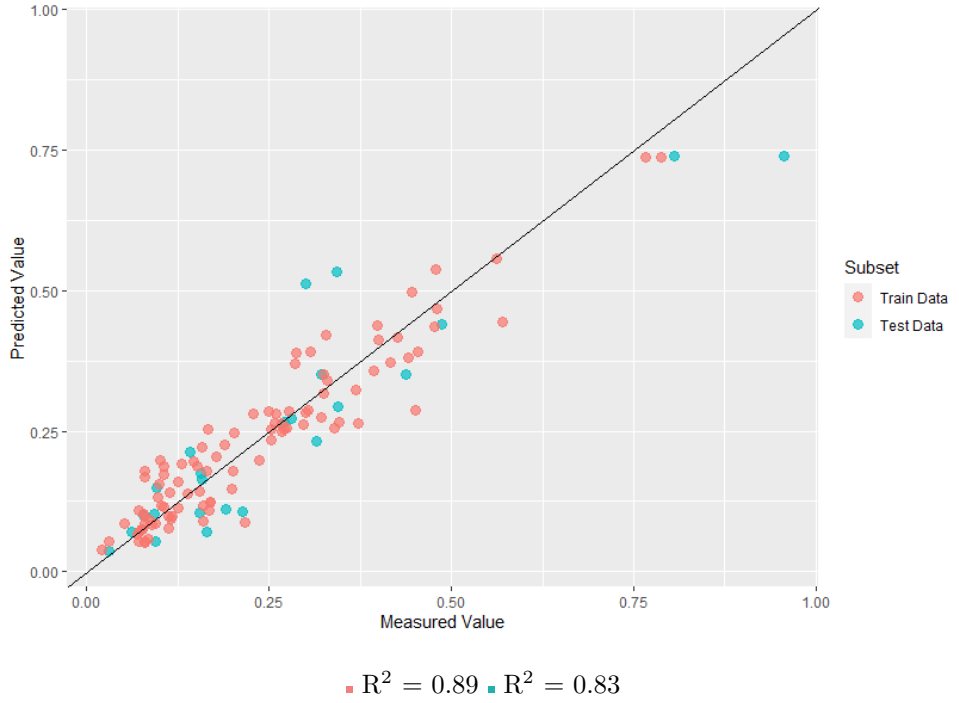


Figure 4.20: Predicted-measured $f_clinear$ for MLR predictive model.

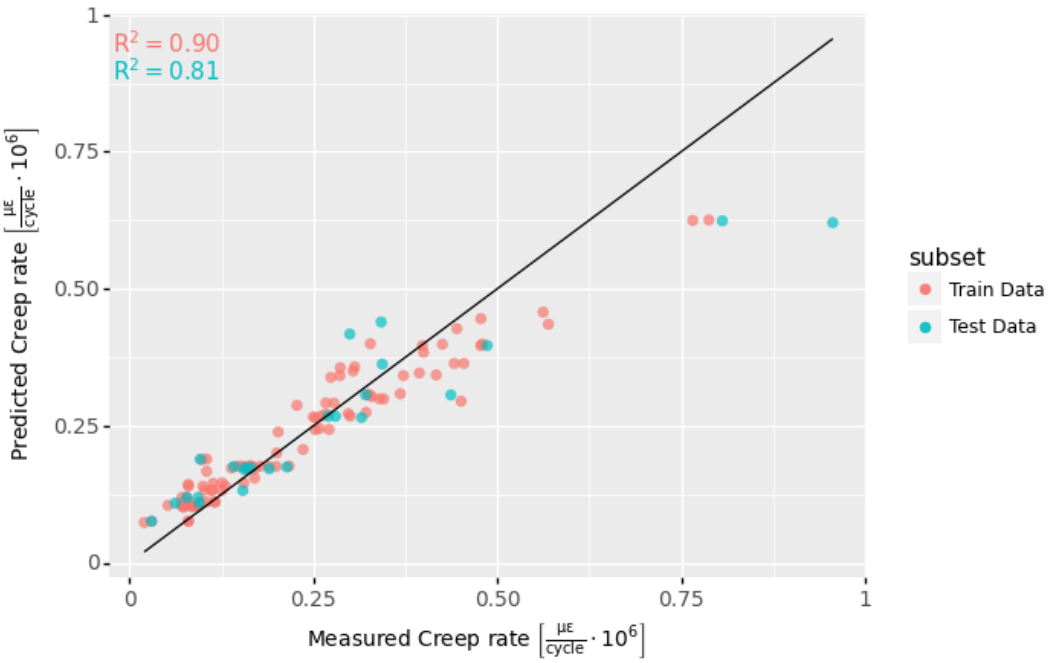


Figure 4.21: Predicted-measured $f_clinear$ for GB predictive model.

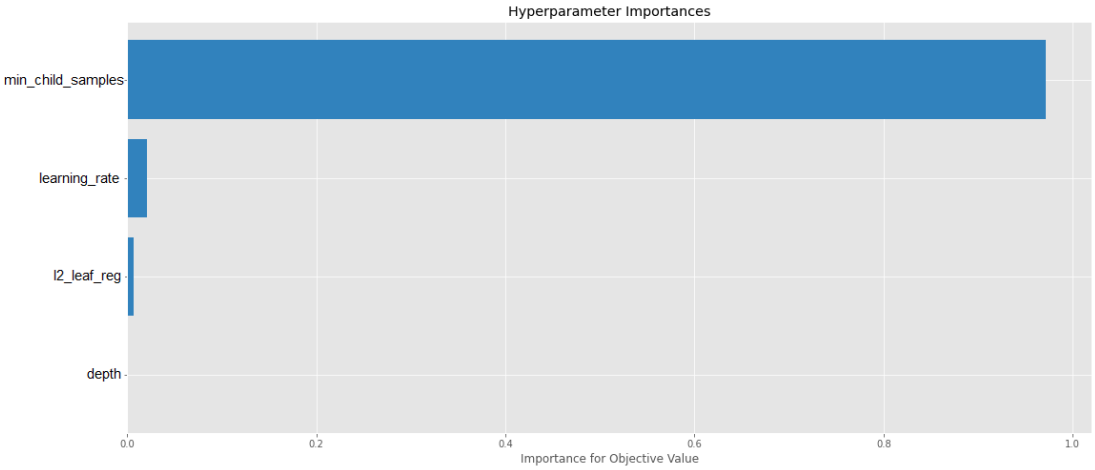


Figure 4.22: f_c linear hyperparameter importance

4.6.3.2 General sensitivity analysis

Resistance to permanent deformation resulted in compaction setup, friction reduction system, and target bitumen mass percentage as the most relevant features. It is important to highlight that the creep rate is the slope of the linear part of the creep curve [31]. The higher the slope, the higher the f_c lin, and the number of cycles to produce a higher strain is reduced. Therefore, the higher the creep rate, the lower the resistance to permanent deformation. The readings from Figure 4.23 show that positive values negatively affected performance.

The compaction setup was found to be the most important feature of the analysis. The gyrator compactor negatively impacted the creep rate, improving the resistance to permanent deformation. However, with a limited number of data points. Segment compactor and field roller positively impacted the creep rate, reducing the resistance to permanent deformation. The field roller had the most data point, resulting in the most reliable conclusions.

The friction reduction system reduces the friction losses in the triaxial test [31]. This system should positively impact the creep rate [15]. Two-layer rubber with silicon grease had the most data points, with the majority positively affecting the creep rate.

Target bitumen content showed that high bitumen percentages increased the creep rate of the mixture. The increase of binder content in the mix increases the susceptibility to rutting [4].

The analysis also resulted in features that disagreed with the findings in the literature review or inconclusive features. The target filler percentage and aging disagreed with expectations. The increase in filler was expected to improve the resistance to permanent deformation. However, it was composed by values lower than 8%, and according to the literature, improvement in resistance to permanent deformation was seen from 10% [46]. The aging is discussed in the

next section. The target composition and RAP showed inconclusive results, possibly caused by the low variability in the data.

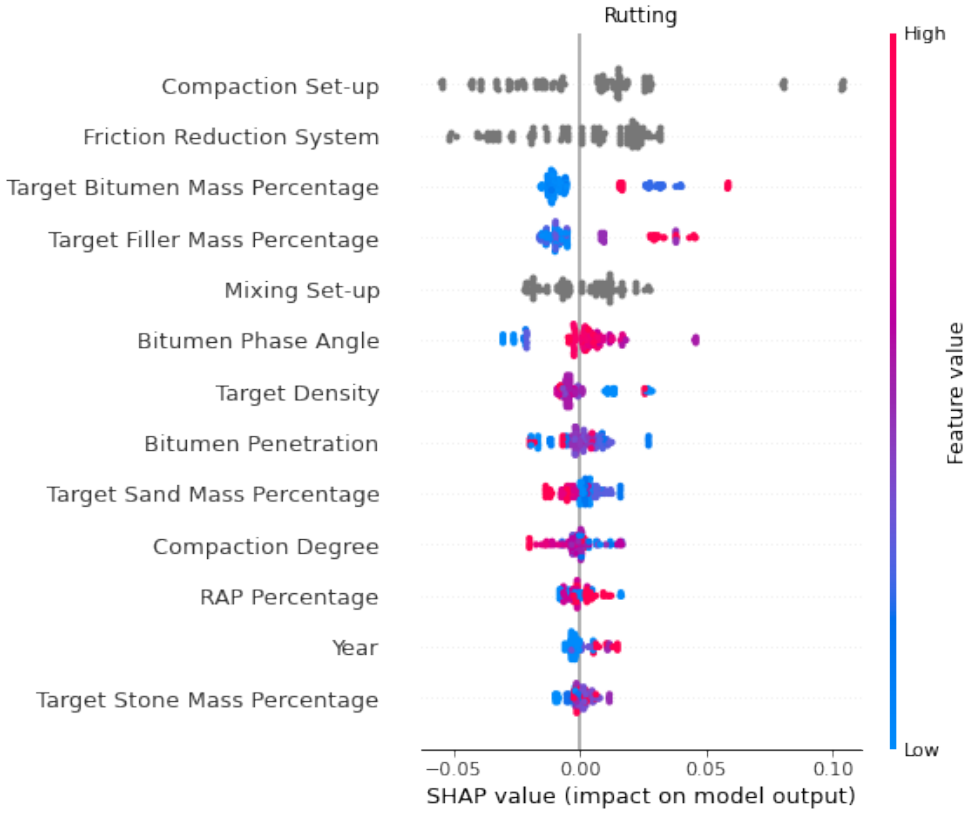


Figure 4.23: Summary SHAP plot

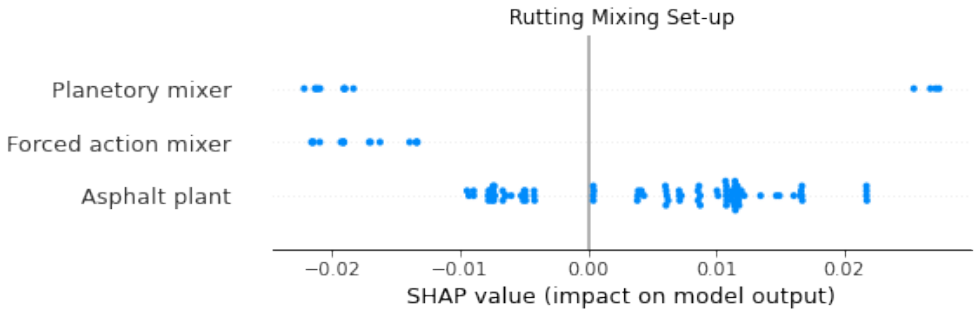


Figure 4.24: Summary SHAP plot for mixing setup

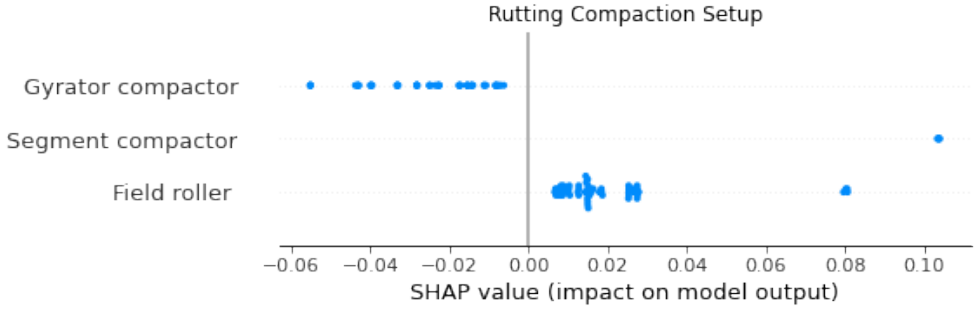


Figure 4.25: Summary SHAP plot for compaction setup

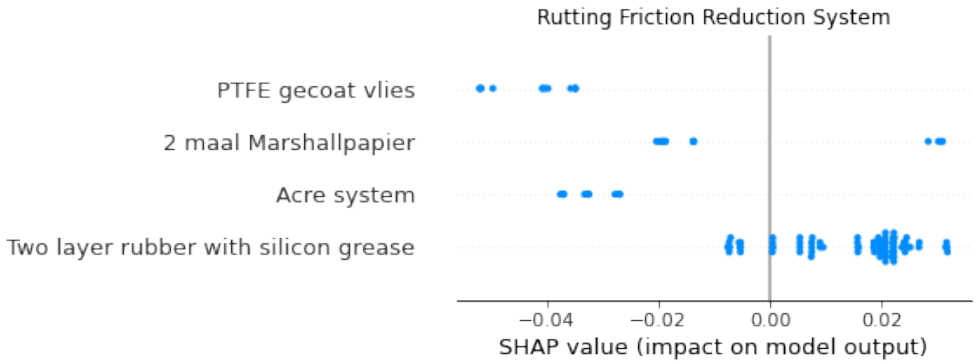


Figure 4.26: Summary SHAP plot for friction reduction system

4.6.3.3 Hypotheses

The functional properties were divided into three main hypotheses determined in the NL-LAB, presented in Section 2.2. Table 4.16 summarizes the hypotheses:

Table 4.16: Resistance to permanent deformation Hypotheses

Hypothesis	Effect on the property
More and softer bitumen	Lower/Smaller
Higher density	Higher/Larger
Aging	Higher/Larger

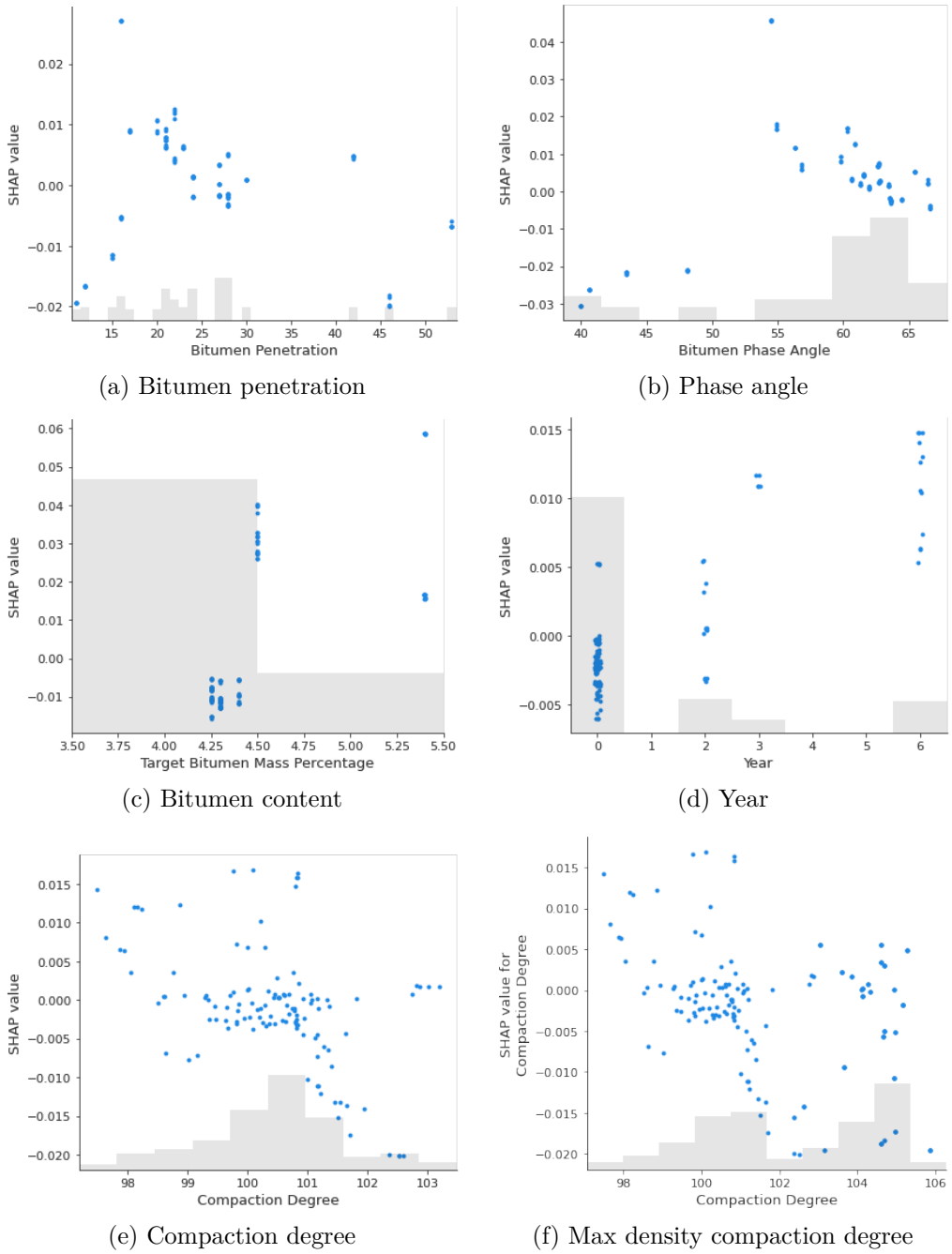
Figure 4.27 a to c show the impact of bitumen content and its tested properties. Figure 4.27 a and b show that increasing the penetration and phase angle reduces the resistance to permanent deformation. Stiffer binders and low phase angles are desirable for rutting prevention [37]. The increase in penetration and phase angle increase the creep rate, reducing the resistance to permanent deformation. However, after a certain point, the positive impact starts to reduce for the increasing values. The bitumen content in Figure 4.27 c shows that higher bitumen content results in lower resistance to permanent deformation [4].

Figure 4.27 d shows that the aging of the material leads to a higher creep rate, hence, lower rutting resistance. Aging leads to a stiffer binder, expecting a higher resistance [37]. It is important to highlight that referring to Figure 4.23, aging was found to have the second lowest overall contribution to the rutting resistance.

Figure 4.27 e and f are plots based on the compaction degree and compaction degree considering maximum density. Higher density results in higher resistance to permanent deformation. However, in design, insufficient compaction can lead to the need for more binder to reach the desired density, and the mix becomes more susceptible to rutting [23].

From a data perspective, the model showed the most results agreeing with the hypotheses. Higher bitumen content and softer bitumen reduce the resistance to permanent deformation.

Higher density leads to lower f_c lin, agreeing with the density hypothesis. The aging results in a higher f_c lin, disagreeing with the hypothesis. The aging results were the opposite of expectations.



(a) Bitumen penetration

(b) Phase angle

(c) Bitumen content

(d) Year

(e) Compaction degree

(f) Max density compaction degree

Figure 4.27: Individual SHAP values

4.6.3.4 Phases analysis

The results from the model regarding resistance to permanent deformation showed that the aging process increases the creep rate. Figure 4.29 shows a trend towards a positive influence in the creep rate, meaning a reduction in rutting resistance. Aging should positively affect rutting resistance because the material becomes stiffer [37]. The Phase difference showed different

ranges for all Phases with a reduction of the resistance against permanent deformation from the laboratory to the field. This indicates differences between the laboratory and the field.

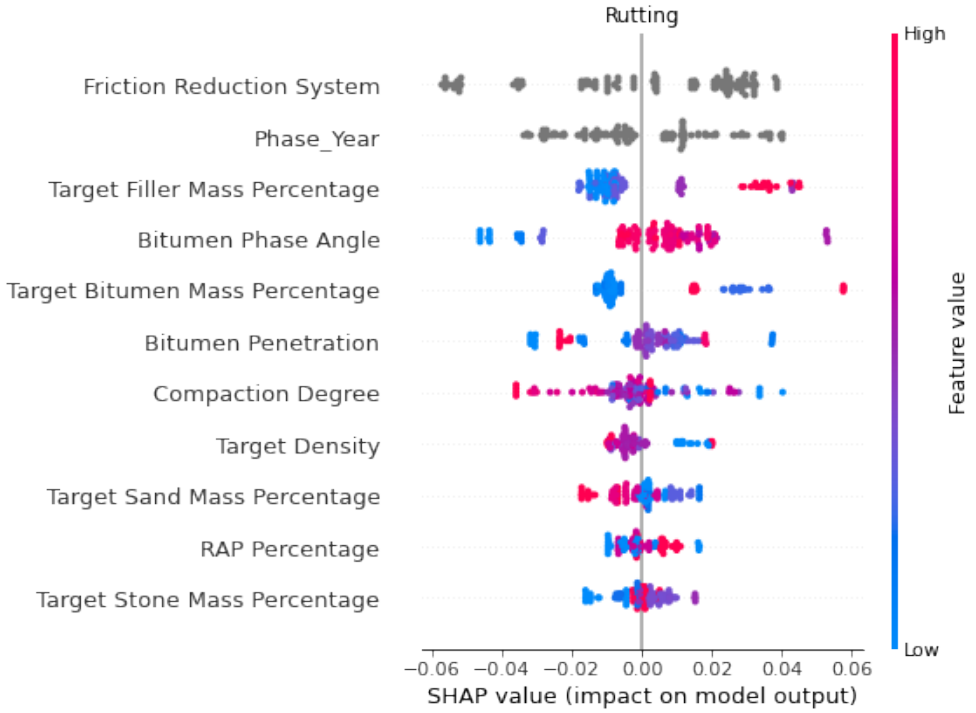


Figure 4.28: Summary SHAP plot

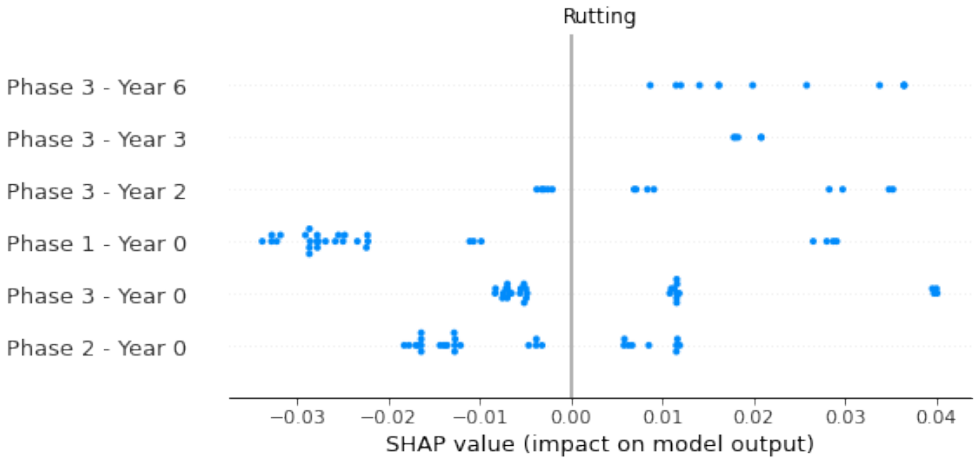


Figure 4.29: Summary SHAP plot for Phase

4.6.4 Indirect tensile strength

4.6.4.1 Predictive performance

The statistical and machine learning models resulted in good prediction accuracy for the ITS, as shown in Figures 4.30 and 4.31. GB outperformed MRL by approximately 10% ($R^2_{Test,GB}$ = 0.84 and $R^2_{Test,MLR}$= 0.75), as shown in Figures 4.30 and 4.31. The machine learning model overestimates the predictions for low ITS values and underestimates values above 2.5 MPa. In

GB, the "min child samples" was found to be the most relevant hyperparameter for the analysis. Followed by the l2 leaf reg, learning rate, and tree depth, presented in Figure 4.32.

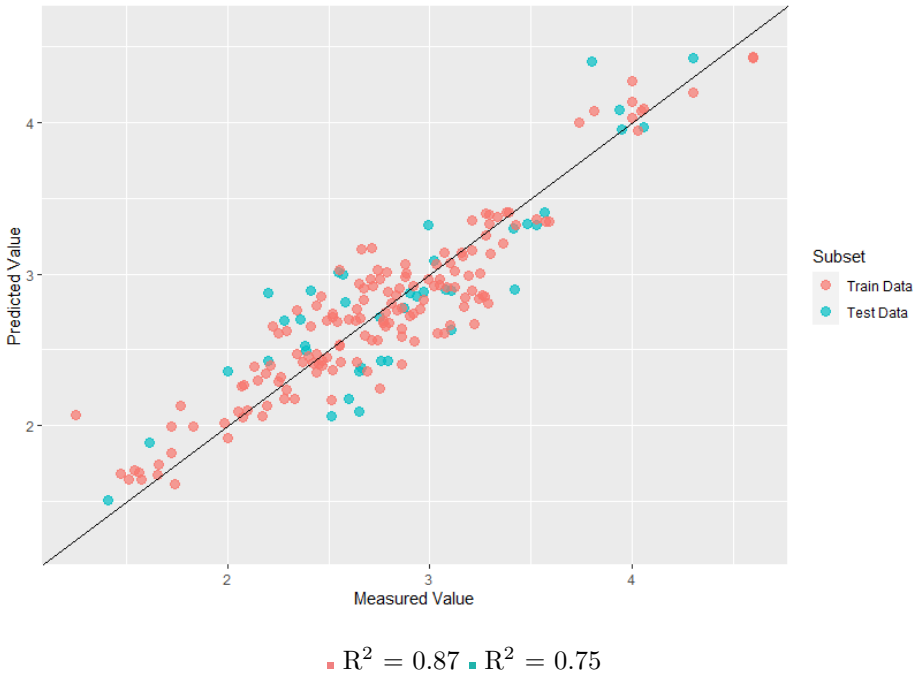


Figure 4.30: Predicted-measured ITS for MLR predictive model.

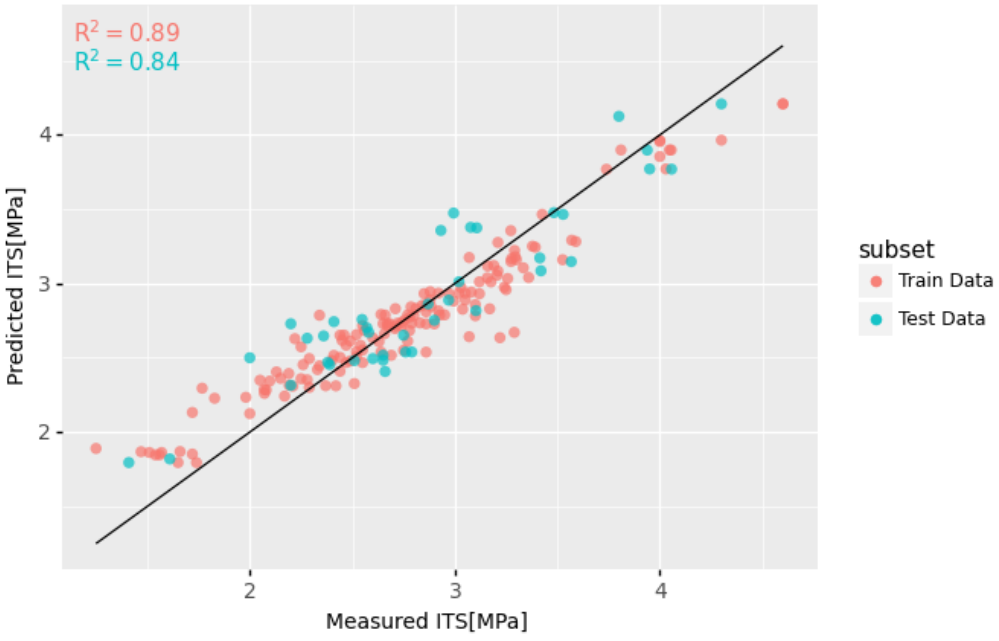


Figure 4.31: Predicted-measured ITS for GB predictive model.

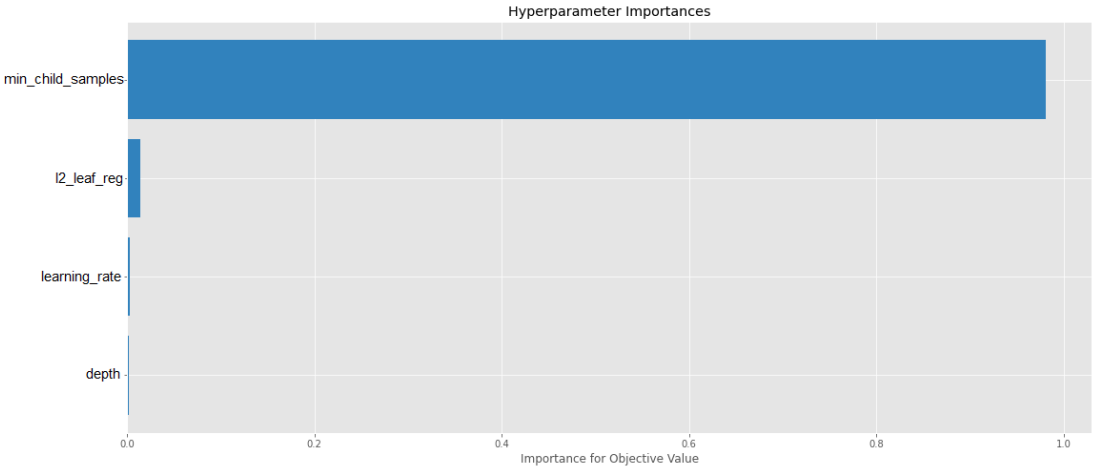


Figure 4.32: ITS hyperparameter importance

4.6.4.2 General sensitivity analysis

The indirect tensile strength model resulted in sample condition, bitumen phase angle, and compaction degree as the most important features. Sample condition defines the sample as wet or dry for the ITS test [34]. Sample condition reduces the binding strength of the bitumen with aggregates, decreasing the ITS [33]. Figure 4.36 shows that wet samples negatively impact the ITS.

Increasing the phase angle reduced the ITS, as seen in Figure 4.33. The high phase angles mean that the material entered the viscous state reducing the elastic behavior.

The compaction degree affects the air void content of the mixture. Low compaction levels result in high void contents leading to cracking and water damage [23]. The increase in density also increases the cohesion of the mix, increasing the ITS.

The results for RAP content differed from expectations. The increase in RAP, according to the literature, increases the ITS [45]. The results for the target sand and stone were inconclusive. Similar to stiffness, fatigue, and rutting, the low variability of the data could be a possible reason for the models not capturing the influence.

The following section aims to deal with ITS and water sensitivity. The ITSR measures water sensitivity, and increasing the ITSR reduces the water sensitivity. The following explanations focus on the phenomena of water sensitivity. Increasing water sensitivity is a negative factor for the mixtures. Hypothesis one and two mean that the ITSR is improved, and hypothesis three means that it has deteriorated.

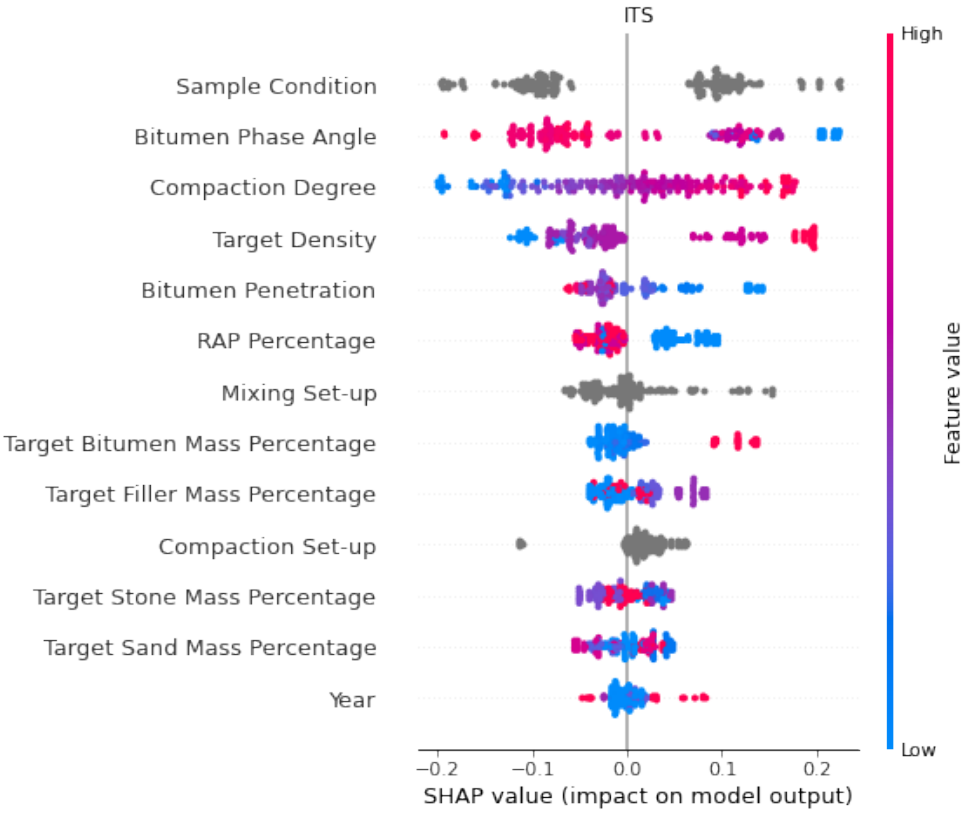


Figure 4.33: Summary SHAP plot

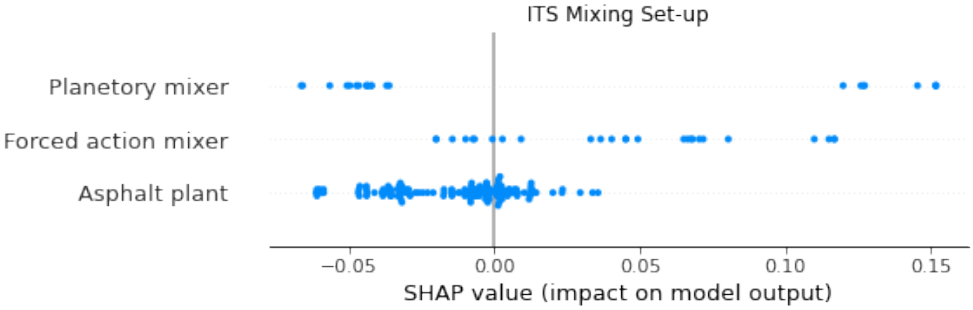


Figure 4.34: Summary SHAP plot for mixing setup

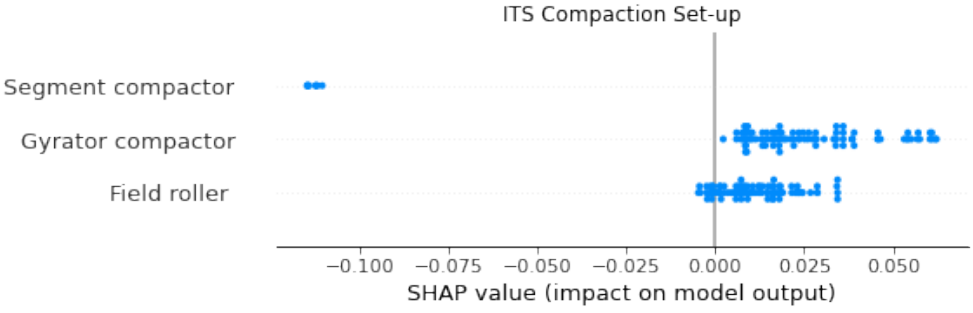


Figure 4.35: Summary SHAP plot for compaction setup

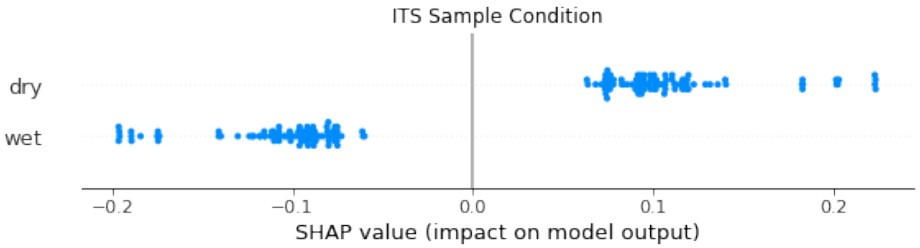


Figure 4.36: Summary SHAP plot for sample condition

4.6.4.3 Hypotheses

The ITS was divided into three main hypotheses determined in the NL-LAB, presented in Section 2.2. Table 4.17 summarizes the hypotheses:

Table 4.17: Water sensitivity Hypotheses

Hypothesis	Effect on the property
More and softer bitumen	Higher/Larger
Higher density	Higher/Larger
Aging	Lower/Smaller

Figure 4.37 a to c shows the bitumen content and its tested properties. The plots were divided into different colors to distinguish the differences between conditioned and unconditioned samples to describe the water sensitivity. Figure 4.37 a and b show that increasing the penetration and phase angle reduce the ITS. On the other hand, the increase of bitumen content, Figure 4.37 c, shows an increase in the ITS. However, it is not clear the difference in the distance between the conditioned and unconditioned data points. A higher binder content reduces the cracking susceptibility of the material, avoiding water damage. Hence, a lower water sensitivity.

The aging of the material in Figure 4.37 d shows inconclusive results for ITS. However, the data points cluster for year 0. By moving toward year six, there was an apparent increase in the distance, indicating a higher water sensitivity. Aging leads to a stiffer binder with higher cracking potential, leading to a higher water sensitivity [43]. It is important to highlight that referring to Figure 4.33, aging was found to have the lowest overall contribution to the ITS. Figures 4.37 e and f are plots based on the compaction degree and compaction degree considering maximum density. Higher density results in higher ITS. For values above 100% compaction degree, the data points grouped, reducing the space, indicating a reduction in the water sensitivity with the increase in density. However, similar to stiffness and fatigue, there were limitations to the positive impact on ITS. For 101.5% compaction degree, the contribution to the ITS reached a maximum, and above, there was no improvement, meaning that, similarly to stiffness and resistance to fatigue, there is a maximum contribution from density to the ITS. Moreover, the SHAP plot also presented an apparent sigmoidal shape, indicating that the resistance would not increase after a limit, remaining constant with the increase in density.

From a data perspective, higher bitumen penetration and phase angles reduce the ITS of the material, and higher content increases the ITS. The ITSR provided inconclusive results. The density hypothesis agrees with the readings showing that higher density leads to higher ITS and lowers water sensitivity. Moreover, similarly to stiffness and resistance to fatigue, the shape found for the compaction degree plot could be potentially used to determine the target density of mixtures. Fitting a sigmoidal function would allow determining the point at which the density stops improving the ITS. However, more tests considering even denser samples should be conducted to understand the impact of extreme densities on the ITS of asphalt mixtures. The aging effect is inconclusive for ITS and ITSR. However, aging is the least important feature in the analysis.

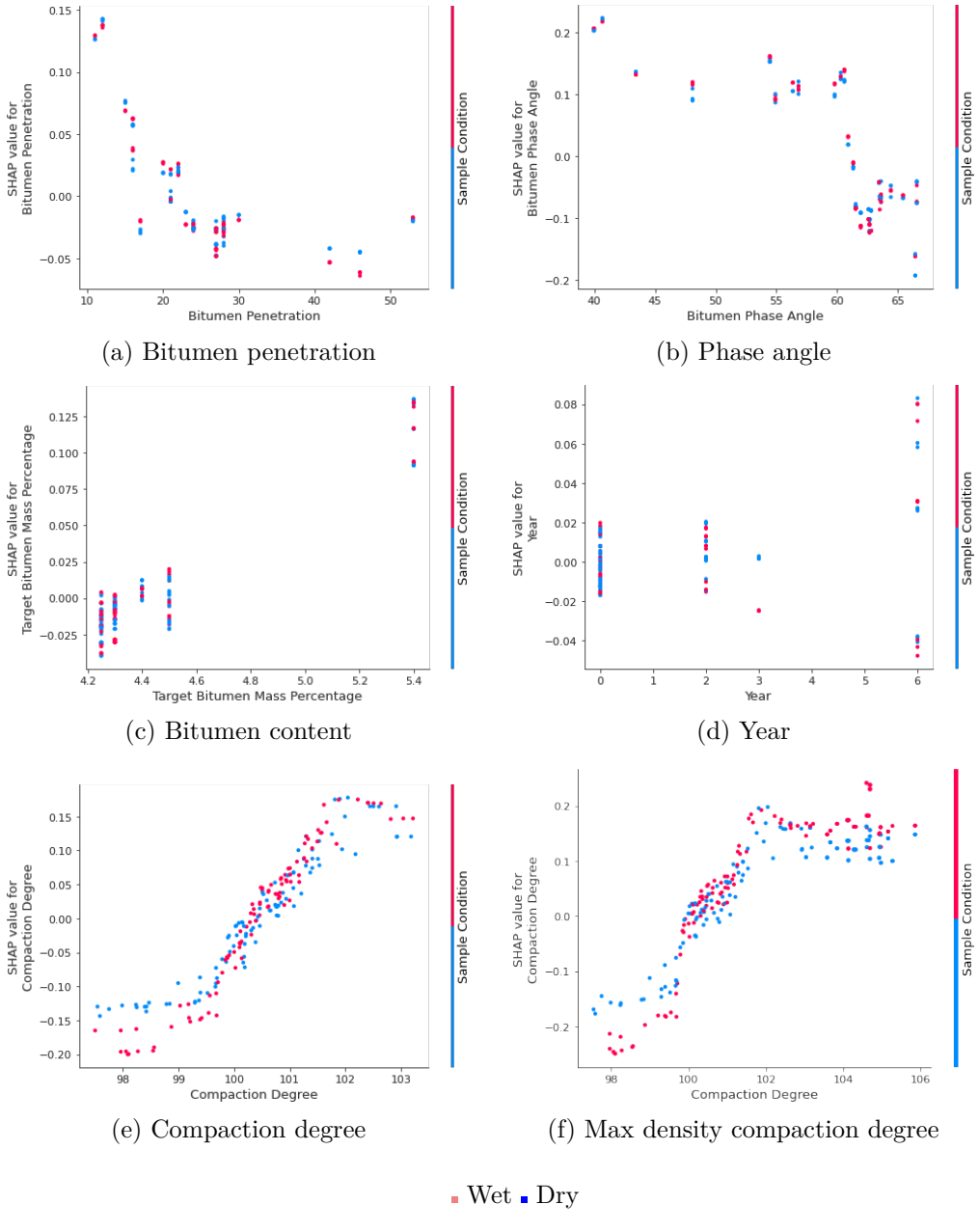


Figure 4.37: Individual SHAP values

4.6.4.4 Phases analysis

The ITS model showed inconclusive results regarding aging. There is no clear trend of improvement or deterioration in Figure 4.38. However, comparing Figures 4.40 and 4.41, the different years of Phase 3 provide similar impacts on the ITS. Moving towards year 3, the negative impact is more significant for wet samples indicating a higher water sensitivity to aged samples. The different Phases show a similar range of influence in the model, and a similar range means that the differences between laboratory and field are not impactful for the ITS.

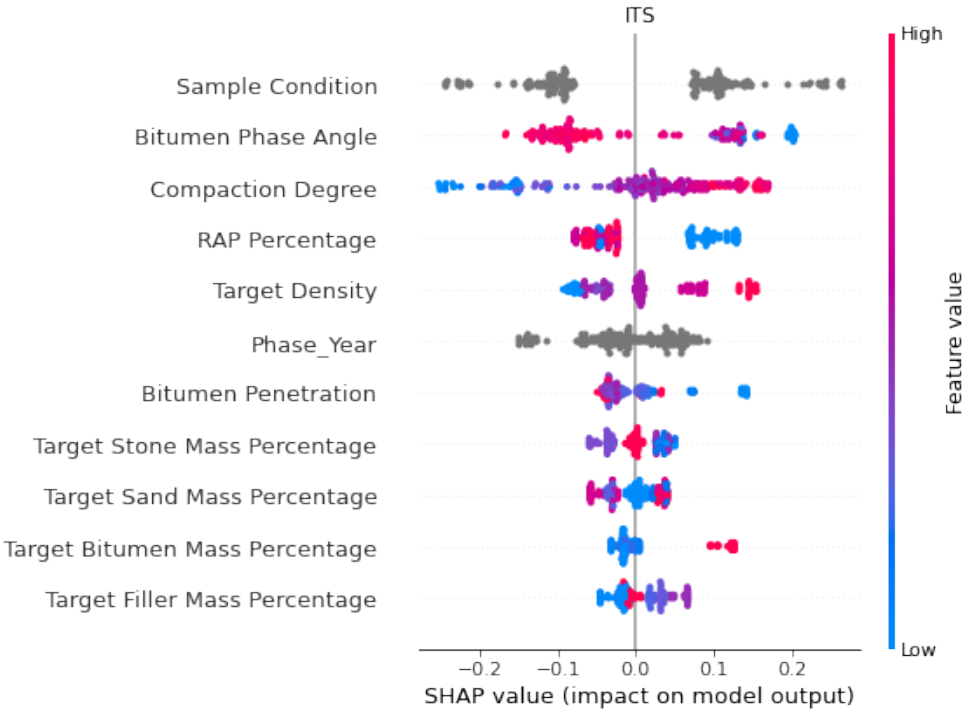


Figure 4.38: Summary SHAP plot

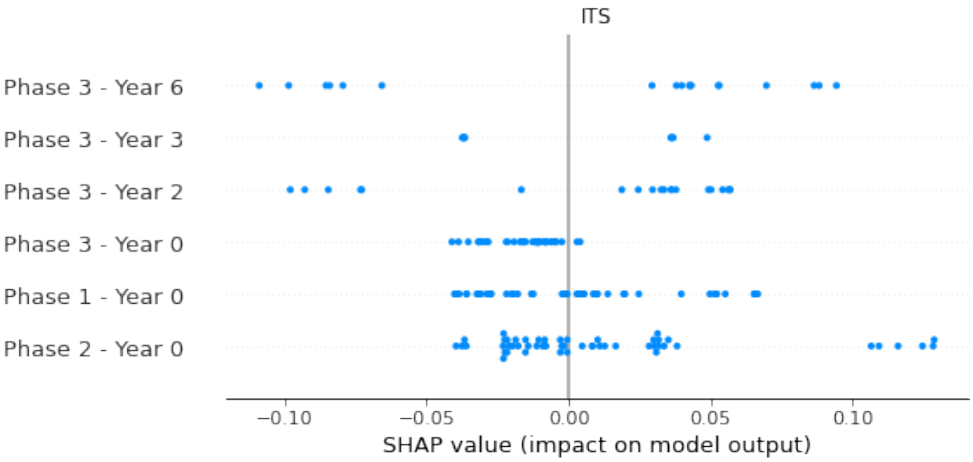


Figure 4.39: Summary SHAP plot for Phase

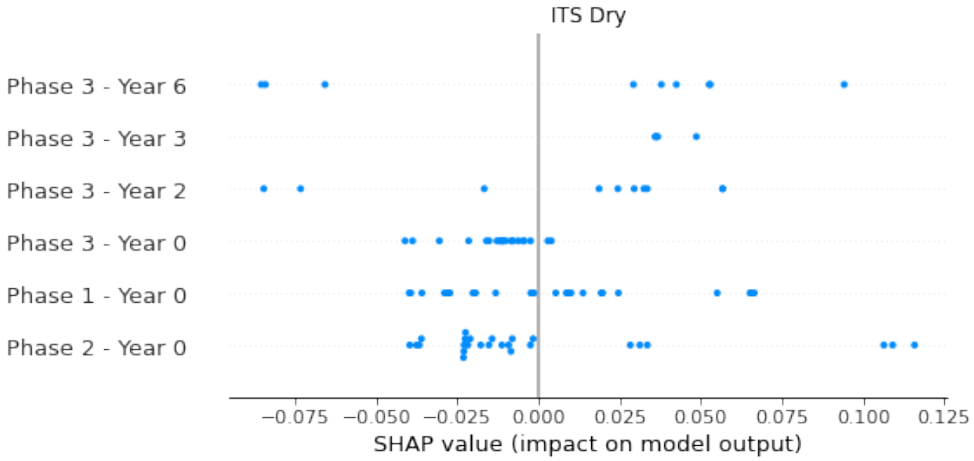


Figure 4.40: Summary SHAP plot for dry samples

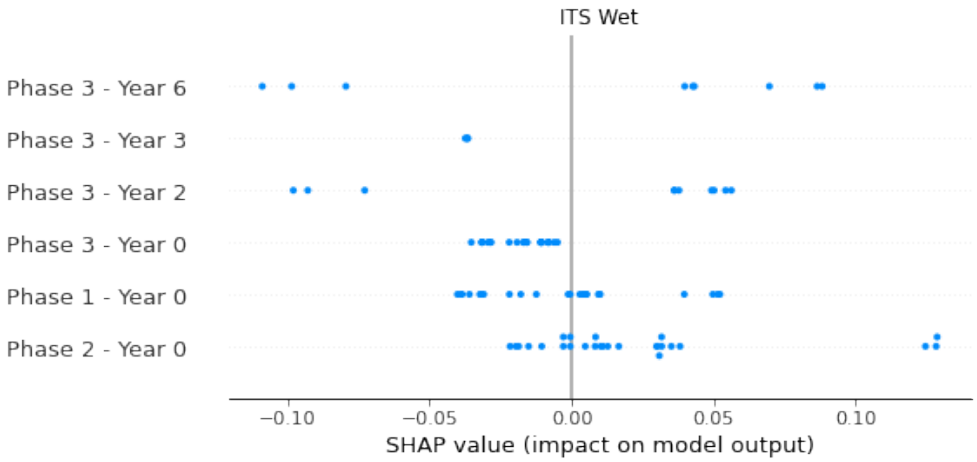


Figure 4.41: Summary SHAP plot for wet samples

4.7 Conclusion

Summarising Chapter 4, the applicability of data-driven approaches has shown positive results for the NL-LAB data set. The summary plots could identify the key parameters for the functional properties of asphalt mixtures, and more attention was given to the three most important parameters. However, some target composition percentages provided inconclusive results, while RAP and aging showed diverging results. The variability of These features was low, and the target composition did not represent the real composition of the mixtures. These issues are possible causes for the model's lack of interpretation.

Regarding the sensitivity analysis focusing on the hypothesis. Most of the results agreed with the hypotheses defined during the NL-LAB project. The remaining results disagreed or were inconclusive, needing further investigation.

The sensitivity analysis that was conducted to distinguish between laboratory and field showed most differences between Phases for fatigue resistance and resistance to permanent deformation. The remaining results were found to be similar, meaning that the differences found in the field and laboratory were minimal.

5

Conclusion and recommendations

This chapter aims to provide the conclusions from the research, followed by recommendations for further research.

5.1 General conclusions

The statistical and machine learning models provided similar results. The similarity between approaches shows that machine learning models are useful tools that can provide reliable predictions. From the author's perspective, both models should be used to support the conclusions and validate each other.

The usefulness of machine learning goes beyond making predictions. The models could be used to help the industry. From the contractor's perspective, the models could be used to better understand the impact of mixing components into the mixture, and to improve the quality of the mixture, providing more sustainable and robust pavement structures. From the researcher's perspective, machine learning models could be used to pinpoint the changes in certain properties on different scales and focus on these properties. The models could also be used to improve/develop type tests.

The framework was able to validate and invalidate the unknown hypotheses developed during the NL-LAB project. Some hypotheses require further research to fully comprehend the impacts of parameters and properties into the mix.

5.2 Conclusion related to the initial hypotheses

Chapter 2 provided an overview of the hypotheses collected during NL-LAB, and Chapter 4 provided the models' results based on the hypotheses. The hypotheses assessed were the following:

Table 5.1: Main Hypotheses

Hypothesis		Effect on the property	Property
1	More and softer bitumen	Lower/Smaller	Stiffness Modulus
		Higher/Larger	Resistance to fatigue
		Lower/Smaller	Resistance to permanent deformation
		Higher/Larger	Water sensitivity
2	Higher density	Higher/Larger	Stiffness Modulus
		Higher/Larger	Resistance to fatigue
		Higher/Larger	Resistance to permanent deformation
		Higher/Larger	Water sensitivity
3	Aging	Higher/Larger	Stiffness Modulus
		Lower/Smaller	Resistance to fatigue
		Higher/Larger	Resistance to permanent deformation
		Lower/Smaller	Water sensitivity

5.2.1 Stiffness

The hypotheses regarding stiffness, presented in Table 5.1, were in agreement with the results found by the model (See Section 4.6.1.3).

(1) The increase in binder content and softer bitumen reduce the stiffness of the mixture. The consistency of the bitumen was found to be of high importance in the model, while the content was less important. The low influence resulted from the small range of binder content used, which caused a low impact on the property.

(2) The increase in density increase the stiffness. The compaction degree was found to be the most important feature of the model. However, there was an apparent limit to the influence of density on stiffness. The SHAP plots, Figure 4.8 show a sigmoidal shape and had no improvement after reaching 101.5% compaction degree.

(3) The aging process increased stiffness. However, the importance was relatively low as compared to compaction degree.

5.2.2 Resistance to fatigue

The resistance to fatigue model shows results in agreement with the hypotheses presented in Table 5.1. However, the bitumen consistency hypotheses showed the opposite behavior (See Section 4.6.2.3.

(1) The bitumen consistency showed an opposite trend as the initial hypothesis. It was found that softer bitumen leads to lower resistance to fatigue. On the other hand, a higher bitumen content resulted in higher resistance to fatigue, agreeing with the hypothesis.

The binder content had the highest importance in the model. Whereas the bitumen consistency had very low importance for penetration and low importance for phase angle.

(2) The increase in density increased the fatigue resistance. However, similarly to stiffness, the plot had a sigmoidal shape, with an apparent limit for a certain threshold. From the scatter plot, values above 101.5% of the compaction degree had no apparent improvement, and the compaction degree had relatively low relevance to resistance to fatigue.

(3) The aging affected fatigue resistance negatively. However, had low importance for the overall model.

5.2.3 Resistance to permanent deformation

The model presented inconclusive results, as shown in Figures 4.23 and 4.27 (See Section 4.6.3.3.

(1) A higher penetration and phase angle values increase the creep rate and lower resistance to permanent deformation. However, a decreasing trend was found for penetration value above 20 and phase angles of 55°, decreasing the creep rate. The bitumen content shows that increasing the mass percentage increases the creep rate, and the quantity of bitumen had a higher importance for the model than phase angle and penetration. This is in agreement with the hypothesis.

(2) A higher density reduces the creep rate, increasing the resistance to permanent deformation. This is in agreement with the hypothesis.

(3) Aging increases the creep rate. This is not in agreement with the hypothesis. However, the aging process had very low importance to the model.

5.2.4 Water sensitivity

The ITS model shows most results in agreement with the hypotheses presented in Table 5.1. However, the bitumen consistency hypotheses showed the opposite behavior (See Section 4.6.4.3.

(1) Softer bitumen leads to lower ITS. The increasing penetration and phase angle did not show a conclusive change in the distance between the data points of wet and dry samples.

(2) The increase in density increases the ITS. Moreover, for values above 100% compaction degree, the data points clustered, and the slope of wet samples was higher than dry samples. The clustering and the increase of slope of wet samples result in an improvement in the ITSR, reducing the water sensitivity. Similarly to stiffness and resistance to fatigue, the improvement in ITS had a sigmoidal shape and had no improvement for values above 102%.

(3) Aging had inconclusive results. However, Figures 4.40 and 4.41 show that conditioned samples had a higher impact on the ITS over the years, indicating that aging reduces the ITSR, increasing the water sensitivity. This is in agreement with the hypothesis.

5.3 Answer to research questions

5.3.1 Answer to main research question

Does the machine learning framework/models developed in this research improve the performance prediction of asphalt mixtures?

Yes, the models improved the performance prediction of asphalt mixtures. The improvement can be seen by evaluating two main aspects:

- Accuracy
- Interpretability

Comparing the accuracy with the classical statistical model, Table 5.5 shows that gradient boosting was outperformed only by the resistance to permanent deformation multiple linear regression model. The remaining machine learning models provided higher accuracies, and by increasing the number of data points, they would become even more robust and most likely provide even better predictions.

The interpretability of the model was the main novelty in the analysis. The feature importance could extract physical interpretation from most of the features of the mixture's properties. The SHAP values, even though no physical meaning was applied, could capture the impact of different features in the analysis and its sensitivity. The hypotheses explained at the beginning of the chapter highlight the interpretability of the model.

5.3.1.1 Comparison to existing machine learning model

The results found in this research could be compared to the machine learning model developed by Martini (2019) because they also used the dataset from NL-Lab [22]. The differences between models were the number of data points, the final selection of features, and the development of a fatigue model. Even though the differences, the models could be compared because of the similar approaches and type of data used. The accuracy of stiffness and resistance to permanent deformation models were higher in Martini's research. However, ITS and resistance to fatigue were higher in this research. The models developed in this research had a more laborious optimization and cross-validation. The cross-validation process resulted in more generalized models. Moreover, the final models used the same mixing composition as Martini. However, condensing the aggregates into group percentages reduced the number of features, reducing the chances of over-fitting and making the models lighter.

Table 5.2: Comparison between machine learning models - 1

Stiffness		ε_6 **Individual**	
Current model	Martini's model	Current model	Martini's model
0.95	0.96	0.8	n/a

Table 5.3: Comparison between machine learning models - 2

f_c lin		ITS	
Current model	Martini's model	Current model	Martini's model
0.81	0.86	0.84	0.82

The interpretability of Martini's model was based on feature importance, where no physical meaning could be read, while with SHAP values, it was possible.

5.3.2 Answer to sub-research questions

How does the machine learning model compare to the statistical model?
The prediction performance of the statistical model and machine learning model was similar. The machine learning model outperformed the statistical model in most of the analyses. The statistical model outperformed only for resistance to permanent deformation. From the author's perspective, the linear model outperformed machine learning for rutting prediction because the creep rate is the linear part of the creep curve, and a linear model could fit the data accurately. Moreover, the limited number of data points for creep rate and ITS decreased the learning performance of the models. By increasing the number of data points, the data's non-linearity would be clearer, and more robust methods perform better than linear models.

The machine learning model had more functionalities for interpreting the results. Multiple linear regression gives the weights of each parameter. However, the machine learning model allowed better visualization by using SHAP plots. Moreover, SHAP plots provide a stronger understanding of the impacts of each data point and feature in the analysis.

The differences in models make both valuable tools for prediction and interpretability. The statistical model can be used for quick and simple predictions. The machine learning model is more robust and can be used for deeper sensitivity analysis and to deal with the increasing complexity of adding more data points.

Table 5.4: Model comparison

	MLR Model		**GB Model**	
Property	**Train**	**Test**	**Train**	**Test**
Stiffness	0.906	0.908	0.954	0.951
ε_6 individual	0.809	0.789	0.83	0.802
f_c lin	0.89	0.833	0.896	0.812
ITS	0.865	0.748	0.892	0.835

Table 5.5: Model comparison overview

Model comparison		
Property	**Train**	**Test**
Stiffness	GB	GB
ε_6 individual	GB	GB
f_c lin	GB	MLR
ITS	GB	GB

What hyperparameters had the most relevance for the analysis? For all models, the most important hyperparameters were min child samples. The hyperparameters optimized in the models were learning rate, depth, l2 leaf reg, and min child samples. The choice of hyperparameters was arbitrary. The number of parameters was limited to four to reduce the computational demand to optimize the models.

What are the optimal hyperparameters after applying different loss functions? The three different loss functions used in the analysis resulted in the hyperparameters in Tables 5.6 and 5.7. For different loss functions, the hyperparameters had slight changes. For the resistance to fatigue model, the quantile loss was unable to finalize the optimization process. Although the change of loss function impacted the hyperparameters' values, the models' accuracy had a minor impact. In general, the selection of the loss function will have more impact on the performance than the selection of hyperparameters. Different loss functions have different optimum hyperparameters that could result in similar accuracy.

Table 5.6: Hyper-parameters for different loss functions - 1

Property	**Stiffness**			**Fatigue**		
Loss Function	RMSE	Quantile	Expectile	RMSE	Quantile	Expectile
Learning rate	0.016	0.018	0.017	0.01	*	0.006
Depth	10	9	9	10		9
l2 leaf reg	2.5	4.5	3.5	1		4
Min child samples	8	32	32	32		32

* The fatigue model did not converge for the optimization utilizing quantile loss.

Table 5.7: Hyper-parameters for different loss functions - 2

Property	**Rutting**			**ITS**		
Loss Function	RMSE	Quantile	Expectile	RMSE	Quantile	Expectile
Learning rate	0.02	0.018	0.014	0.009	0.017	0.003
Depth	13	10	13	11	15	10
Depth	13	10	13	11	15	10
l2 leaf reg	1	3.5	2.5	4	4.5	3.5
Min child samples	1	4	16	32	16	32

Does the machine learning model capture the physical behavior of the mixture? Yes. The machine learning models could capture the mixtures' physical behavior through sensitivity analysis. Parameters with a high correlation with the properties of the asphalt mixture were ranked highly in the analysis. The knowledge about the impacts of mixture

components could be recognized. Some were in line with expectations, and some disagreed.

5.4 Recommendations

The recommendations were divided into two subsets. The first subset relates to the mixture's components and laboratory tests that had unexpected behavior. The second is recommendations for the application of machine learning models.

Material level:

(1) Perform fatigue tests for asphalt mixtures of the same composition, varying only the bitumen's penetration and phase angle. The results found in this thesis diverge from the expectations collected from NL-LAB, and further testing is required. The expectations were that a softer binder would lead to higher fatigue resistance. Collect the data for later application to machine learning tools.

(2) Perform tests to define the resistance to permanent deformation on more samples with different aging. The results found in this thesis diverge from the expectations collected from NL-LAB, and further testing is required. The expectations were that aging leads to higher resistance to permanent deformation. Collect the data for later application to machine learning tools.

(3) Perform tests to define the stiffness, resistance to permanent deformation, and ITS varying the reclaimed asphalt content. The results found in this thesis diverge from the expectations collected from NL-LAB, and further testing is required. The expectations were that increasing RAP content would lead to higher stiffness, resistance to permanent deformation, and ITS. Collect the data for later application to machine learning tools.

(4) Stiffness, fatigue resistance, and indirect tensile strength have shown a sigmoidal shape for the SHAP plots. Testing samples of low and extreme densities to evaluate the importance of density and if the sigmoidal shape continues or if another trend arises. Potentially, a method to define the target density could be defined by fitting a sigmoidal curve for the compaction degree plots to determine the point at which the density does not affect the property anymore. Collect the data for later apply to machine learning tools.

Machine learning level:

(1) Application of the developed models to different datasets to verify the performance for different compositions. In addition, apply continuous learning for the developed models to understand the models' behavior for adding more data.

(2) Development of a user interface for training the models, and the continuation of the training could be applied. This user interface would allow each contractor to train the models with their mixtures. Different models could be trained for single mixtures or multiple mixtures. Models trained for single mixtures could be used to predict the

functional properties, and a deep sensitivity analysis could be applied to improve the mixture's quality. For multiple mixtures, the models could be used to understand the connection between the composition and the functional properties of asphalt concrete.

(3) Development of a physics-informed model. Physics-informed models apply prior knowledge to improve the training of neural networks and reduce the number of data points required.

(4) Develop a machine learning framework for resistance to fatigue utilizing the number of cycles to reach 50% of the initial stiffness. The current models used strain levels extrapolated from the fatigue line, which is unadvised [22]. The proposed model does not require extrapolated data.

(5) Develop a machine learning framework for resistance to permanent deformation utilizing the rutting factor, $G^*/\sin\delta$, of the sample or the accumulated strain for 10000 cycles, ϵ_{10000}. These parameters are continuous values that do not originate from data fitting.

(6) Expansion of NL-LAB dataset for further improvement of the models and studies to be conducted.

(7) Development of frameworks utilizing different types of asphalt concrete mixtures to verify the applicability of machine learning tools.

(8) Switch the target composition to extracted composition in the training process. Models trained on the real composition could better interpret the data and reduce the low variability issue.

(9) Development of data frame with non-conventional materials to gather information for later develop models in the early stages of material usage.

References

[1] SMJG Erkens, Dave van Vliet, and Martin van de Ven. "On the need for innovation in road engineering A Dutch example". In: *3rd international symposium on asphalt pavements and environment, Sun City, South Africa* (2015).

[2] NationMaster. *Countries Compared*. URL: https://www.nationmaster.com/country-info/stats/Transport/Road-density/Km-of-road-per-100-sq.-km-of-land-area (visited on 12/02/2022).

[3] TNO. *Asphalt in The Netherlands*. URL: https://www.tno.nl/en/sustainable/safe-sustainable-living-environment/infrastructure/asphalt/ (visited on 11/09/2022).

[4] Sandra Erkens et al. "NL-LAB: onderzoek naar de voorspellende waarde van proef 62". In: *CROW Infradagen* (2014), pp. 18–19.

[5] Paulo Pereira and Jorge Pais. "Main flexible pavement and mix design methods in Europe and challenges for the development of an European method". In: *Journal of Traffic and Transportation Engineering (English Edition)* (2017). DOI: http://dx.doi.org/10.1016/j.jtte.2017.06.001.

[6] Qiao Dong et al. "Data Analysis in Pavement Engineering: An Overview". In: *IEEE Transactions on Intelligent Transportation Systems* (2021). ISSN: 15580016. DOI: 10.1109/TITS.2021.3115792.

[7] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).

[8] Eva Ivanova and Jana Masarova. "Importance of road infrasctructure in the economic development and competitiveness". In: *Economics and Management:2013.18(2)* (2013). DOI: http://dx.doi.org/10.5755/j01.em.18.2.4253.

[9] Robert Puentes. *Why Infrastructure Matters: Rotten Roads, Bum Economy*. 2015. URL: https://www.brookings.edu/opinions/why-infrastructure-matters-rotten-roads-bum-economy (visited on 08/05/2022).

[10] Rijkswaterstaat Ministry of Infrastructure and Water Management. *Roads and waterways*. URL: https://www.rijkswaterstaat.nl/en/mobility/roads-and-waterways (visited on 08/05/2022).

[11] WorldData. *Netherlands*. URL: https://www.worlddata.info/europe/netherlands/index.php (visited on 08/05/2022).

[12] J Th Van der Zwan et al. “Porous asphalt wearing courses in the Netherlands a state of the art review”. In: *69th Annual Meeting of theTransportation Research Board* (1990).

[13] *CROW (2010). Standaard RAW Bepalingen 2010. Standard, CROW.*

[14] Sandra Erkens et al. *NL-LAB 2012 - 2016 - data overview to date* -. 2017.

[15] Georgios Seleridis. *Evaluation Of The Current Test Methods Of Water Sensitivity And Permanent Deformation.* 2017.

[16] *NEN-EN 12697-26:2018 - Bituminous mixtures - Test methods - Part 26: Stiffness. Standard, Techinical Commitee CEN/TC 227, Brussel.*

[17] Y.H.Huang. *Pavement Analysis and Design.* Second Edition. 2004: Prentice Hall, 2017.

[18] Joost Droogers. *Asphalt concrete stiffness prediction based on composition and binder properties.* 2018.

[19] Shell International Petroleum Company. *Shell Pavement Design Manual: Asphalt Pavements and Overlays for Road Traffic.* Shell International Petroleum Company, 1978.

[20] Javed Bari. “Development of a new revised version of the Witczak E Predictive Model for hot mix asphalt mixtures”. In: (2006). URL: https://www.researchgate.net/publication/273216159.

[21] Hong Zhang et al. “Comparison of different micromechanical models for predicting the effective properties of open graded mixes”. In: *Transportation Research Record* 2672.28 (2018), pp. 404–415.

[22] Giulia Martini. *Predictive Modelling of Asphalt Concrete Functional Properties Using Multiple Linear Regression and Gradient Boosting.* 2019.

[23] E Brown. “Density of asphalt concrete-how much is needed?” In: *National Center for Asphalt Technology (US)* (1990). DOI: https://doi.org/10.21949/1404494.

[24] Walaa S Mogawer et al. “Evaluation of the effects of hot mix asphalt density on mixture fatigue performance, rutting performance and MEPDG distress predictions”. In: *International Journal of Pavement Engineering* 12 (2 Apr. 2011), pp. 161–175. ISSN: 10298436. DOI: 10.1080/10298436.2010.546857.

[25] Eliana Del Pilar Vivar and John E Haddock. “HMA pavement performance and durability”. In: *FHWA/IN/JTRP-2005/14* (2006).

[26] E Ray Brown and Stephen A Cross. “A study of in-place rutting of asphalt pavements”. In: *National Center for Asphalt Technology* (1989). URL: https://rosap.ntl.bts.gov/view/dot/14139.

[27] Fernando Moreno-Navarro et al. “Análisis de la susceptibilidad al agua en mezclas bituminosas en caliente mediante el estudio comparativo de dos métodos de ensayo de laboratorio”. In: *DYNA (Colombia)* 81 (183 2014), pp. 49–59. ISSN: 00127353. DOI: 10.15446/dyna.v81n183.30893.

[28] *NEN-EN 12697-24:2018 - Bituminous mixtures - Test methods - Part 24: Resistance to fatigue. Standard, Techinical Commitee CEN/TC 227, Brussel.*

[29] F Bonnaure, A Gravois, and J Udron. "A new method for predicting the fatigue life characteristics of bituminous mixtures". In: *Proceedings of association of Asphalt Paving Technologies (AAPT)* 49 (1980).

[30] Regis Luis Egual De Carvalho. *Prediction of permanent deformation in asphalt concrete.* University of Maryland, College Park, 2012.

[31] *NEN-EN 12697-25:2016 - Bituminous mixtures - Test methods - Part 25: Cyclic compression test. Standard, Techinical Commitee CEN/TC 227, Brussel.*

[32] Sandra Erkens et al. "Comparing lab to field properties in CEN Type Testing for Asphalt Concrete-the NL-LAB program". In: *Proceedings of 7th Transport Research Arena TRA* (2018).

[33] A O Abd El Halim and Manthan Ramani. "Stripping Distress on Hot Mixed Asphalt Pavement". In: *GRD Journals | Global Research and Development Journal for Engineering | Recent Advances in Civil Engineering for Global Sustainability |* (2016).

[34] *NEN-EN 12697-23:2017 -Bituminous mixtures - Test methods - Part 23: Determination of the indirect tensile strength of bituminous specimens. Standard, Techinical Commitee CEN/TC 227 , Brussel.*

[35] *NEN-EN 12697-12:2018 - Bituminous mixtures - Test methods - Part 12: Determination of the water sensitivity of bituminous specimens. Standard, Techinical Commitee CEN/TC 227 , Brussel.*

[36] *NEN-EN 1426:2015 - Bitumen and bituminous binders - Determination of needle penetration . Standard, Techinical Commitee CEN/TC 336, Brussel.*

[37] A Thomas Papagiannakis and Eyad A Masad. *Pavement design and materials.* John Wiley & Sons, 2008, p. 542. ISBN: 9780471214618.

[38] *NEN-EN 1427:2015 - Bitumen and bituminous binders - Determination of the softening point - Ring and Ball method. Standard, Techinical Commitee CEN/TC 336, Brussel.*

[39] *NEN-EN 14770:2012 -Bitumen and bituminous binders - Determination of complex shear modulus and phase angle - Dynamic Shear Rheometer (DSR). Standard, Techinical Commitee CEN/TC 336, Brussel.*

[40] Jeroen Besamusca. *Lecture Notes - 1. Asphalt production and recycling trends. 2. Low temperature asphalt with recycling.* Kuwait Petroleum Research Technology, 2021.

[41] Gebert van Bochove. *Lecture Notes - Road Construction Maintenance - Bitumen.* Heijmans, 2021.

[42] Cheolmin Baek, B Shane Underwood, and Y Richard Kim. "Effects of oxidative aging on asphalt mixture properties". In: *Transportation Research Record* (2296 Dec. 2012), pp. 77–85. ISSN: 03611981. DOI: 10.3141/2296-08.

[43] José P Aguiar-Moya et al. “Effect of aging on adhesion properties of asphalt mixtures with the use of bitumen bond strength and surface energy measurement tests”. In: *Transportation Research Record* 2505 (2015), pp. 57–65. ISSN: 03611981. DOI: 10.3141/2505-08.

[44] Luca Noferini et al. “Investigation on performances of asphalt mixtures made with Reclaimed Asphalt Pavement: Effects of interaction between virgin and RAP bitumen”. In: *International Journal of Pavement Research and Technology* 10 (4 July 2017), pp. 322–332. ISSN: 19971400. DOI: 10.1016/j.ijprt.2017.03.011.

[45] Imad L Al-Qadi et al. “Impact of high RAP contents on structural and performance properties of asphalt mixtures”. In: *Illinois Center for Transportation* (2012). ISSN: 0197-9191.

[46] Baoshan Huang, Xiang Shu, and Xingwei Chen. “Effects of mineral fillers on hot-mix asphalt laboratory-measured properties”. In: *International Journal of Pavement Engineering* 8 (1 Mar. 2007), pp. 1–9. ISSN: 10298436. DOI: 10.1080/10298430600819170.

[47] Min-Chih Liao, Jian-Shiuh Chen, and Ko-Wan Tsou. “Fatigue Characteristics of Bitumen-Filler Mastics and Asphalt Mixtures”. In: *Journal of Materials in Civil Engineering* 24 (7 July 2012), pp. 916–923. ISSN: 0899-1561. DOI: 10.1061/(asce)mt.1943-5533.0000450.

[48] Cambridge Dictionary. *Definition of data-driven*. URL: https://dictionary.cambridge.org/dictionary/english/data-driven (visited on 11/16/2022).

[49] Bradley Efron and Trevor Hastie. *Computer Age Statistical Inference, Student Edition: Algorithms, Evidence, and Data Science*. Vol. 6. Cambridge University Press, 2021.

[50] Brian S Everitt and Anders Skrondal. *The Cambridge Dictionary of statistics*. Fourth Edition. Cambridge University Press, 2002.

[51] Stuart Russell and Peter Norvig. “Artificial Intelligence A Modern Approach”. In: (1994).

[52] Siddhartha Ray Barua. *A Strategic Perspective on the Commercialization of Artificial Intelligence: A socio-technical analysis*. Massachusetts Institute of Technology, 2019.

[53] Probyto Data Science and Consulting Pvt. ltd. *Data science for business professionals: A practical guide for biginners*. Bpb, 2020, p. 542.

[54] L.Nguyen. *EDA, Data Preprocessing, Feature Engineering: We are different!* 2021. URL: https://medium.com/@ndleah/eda-data-preprocessing-feature-engineering-we-are-different-d2a5fa09f527#:~:text=Alternatively\%2C\%20feature\%20engineering\%20is\%20the,scientist\%20to\%20begin\%20creating\%20those (visited on 11/19/2022).

[55] Heavy.AI. *Feature Engineering*. URL: https://www.heavy.ai/technical-glossary/feature-engineering (visited on 11/19/2022).

[56] Mervyn Stone. “Cross-validatory choice and assessment of statistical predictions”. In: *Journal of the royal statistical society: Series B (Methodological)* 36 (2 1974), pp. 111–133.

[57] Kedar Potdar, Taher S Pardawala, and Chinmay D Pai. "A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers". In: *International Journal of Computer Applications (0975 8887) Volume 175 No.4* (2017).

[58] Damodar Gujarati. "Use of dummy variables in testing for equality between sets of coefficients in linear regressions: A generalization". In: *American Statistician* 24 (5 1970), pp. 18–22. ISSN: 15372731. DOI: 10.1080/00031305.1970.10477220.

[59] Patricio Cerda and Gaël Varoquaux. "Encoding High-Cardinality String Categorical Variables". In: *IEEE Transactions on Knowledge and Data Engineering* 34 (3 Mar. 2022), pp. 1164–1176. ISSN: 15582191. DOI: 10.1109/TKDE.2020.2992529.

[60] Liudmila Prokhorenkova et al. "CatBoost: unbiased boosting with categorical features". In: *Advances in neural information processing systems* 31 (2018).

[61] Yandex. *Transforming categorical features to numerical features*. URL: https://catboost.ai/en/docs/concepts/algorithm-main-stages_cat-to-numberic (visited on 10/28/2022).

[62] Vikas Chaurasia and Saurabh Pal. "Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer". In: *SN Computer Science* 1 (5 Sept. 2020). ISSN: 26618907. DOI: 10.1007/s42979-020-00296-8.

[63] scikit-learn developers (BSD License). *1.4 Support Vector Machines*. 2007 - 2022. URL: https://scikit-learn.org/stable/modules/svm.html (visited on 10/15/2022).

[64] Marko Sarstedt, Erik Mooi, et al. *A Concise Guide to Market Research*. Vol. 12. Springer Berlin Heidelberg, 2014. ISBN: 978-3-662-56706-7. DOI: 10.1007/978-3-662-56707-4. URL: http://link.springer.com/10.1007/978-3-662-56707-4.

[65] Marcus Gallagher, Nour Moustafa, and Erandi Lakshika. *AI 2020: Advances in Artificial Intelligence: 33rd Australasian Joint Conference, AI 2020, Canberra, ACT, Australia, November 29–30, 2020, Proceedings*. Vol. 12576. Springer Nature, 2020.

[66] Benyamin Ghojogh and Mark Crowley. "The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial". In: (2019). DOI: 10.48550/ARXIV.1905.12787. URL: https://arxiv.org/abs/1905.12787.

[67] Alan O Sykes. "An Introduction to Regression Analysis". In: *(Coase-Sandor Institute for Law Economics Working Paper No. 20)* (1993). URL: https://chicagounbound.uchicago.edu/law_and_economics.

[68] Kasthurirangan Gopalakrishnan and Sunghwan Kim. "Support Vector Machines Approach to HMA Stiffness Prediction". In: (2011). DOI: 10.1061/ASCEEM.1943-7889.0000214.

[69] Dana Daneshvar and Ali Behnood. "Estimation of the dynamic modulus of asphalt concretes using random forests algorithm". In: *International Journal of Pavement Engineering* 23 (2 2022), pp. 250–260. ISSN: 1477268X. DOI: 10.1080/10298436.2020.1741587.

[70] Jia Wu, SenPeng Chen, and XiYuan Liu. "Efficient hyperparameter optimization through model-based reinforcement learning". In: *Neurocomputing* 409 (Oct. 2020), pp. 381–393. ISSN: 18728286. DOI: 10.1016/j.neucom.2020.06.064.

[71] Li Yang and Abdallah Shami. "On hyperparameter optimization of machine learning algorithms: Theory and practice". In: *Neurocomputing* 415 (Nov. 2020), pp. 295–316. ISSN: 18728286. DOI: 10.1016/j.neucom.2020.07.061.

[72] James Bergstra, Daniel Yamins, and David Cox. "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures". In: *International conference on machine learning*. PMLR. 2013, pp. 115–123.

[73] James Bergstra et al. "Algorithms for hyper-parameter optimization". In: *Advances in neural information processing systems* 24 (2011).

[74] Qi Wang et al. "A Comprehensive Survey of Loss Functions in Machine Learning". In: *Annals of Data Science* 9 (2 Apr. 2022), pp. 187–212. ISSN: 21985812. DOI: 10.1007/s40745-020-00253-5.

[75] Timothy O Hodson. "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not". In: *Geoscientific Model Development* 15 (14 July 2022), pp. 5481–5487. ISSN: 19919603. DOI: 10.5194/gmd-15-5481-2022.

[76] Linda Schulze Waltrup et al. In: *Statistical Modelling* (5), pp. 433–456. ISSN: 14770342. DOI: 10.1177/1471082X14561155.

[77] Sanjay Yadav and Sanyam Shukla. "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification". In: Institute of Electrical and Electronics Engineers Inc., Aug. 2016, pp. 78–83. ISBN: 9781467382861. DOI: 10.1109/IACC.2016.25.

[78] Lloyd S Shapley. "Notes on the n-Person Game – II: The Value of an n-Person Game". In: *RAND Corporation* (Aug. 1951). URL: https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM670.pdf.

[79] S.Lundberg. *SHAP*. 2018. URL: https://github.com/slundberg/shap (visited on 11/04/2022).

[80] Alex J Smola and Bernhard Schölkopf. "A tutorial on support vector regression". In: *Statistics and computing* 14.3 (2004), pp. 199–222.

[81] Mariette Awad and Rahul Khanna. *Support vector regression*. Springer, 2015, pp. 67–80.

[82] David Duvenaud. *The Kernel Cookbook:Advice on Covariance functions*. 2014. URL: https://www.cs.toronto.edu/~duvenaud/cookbook/index.html (visited on 10/25/2022).

[83] Wayback Machine. *Radial functions*. 2014. URL: https://web.archive.org/web/20140424004957/http://www.anc.ed.ac.uk/rbf/intro/node7.html (visited on 11/17/2022).

[84] scikit-learn developers (BSD License). *1.10. Decision Treesű*. URL: https://scikit-learn.org/stable/modules/tree.html (visited on 10/25/2022).

[85] Tin Kam Ho. “Random decision forests”. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition* (1995). DOI: 10.1109/ICDAR.1995.598994.

[86] Robert E Schapire. “The strength of weak learnability”. In: *Machine learning* 5.2 (1990), pp. 197–227.

[87] Michael Kearns and Leslie Valiant. “Learning Boolean Formulae or Finite Automata is as Hard as Factoring”. In: TR 14-88 (1988).

[88] Runhua Guo, Donglei Fu, and Giuseppe Sollazzo. “An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree”. In: *International Journal of Pavement Engineering* 23 (10 2021), pp. 1–14. ISSN: 1477268X. DOI: 10.1080/10298436.2021.1910825.

[89] scikit-learn developers (BSD License). *1.11.4. Gradient Tree Boostingű*. URL: https://scikit-learn.org/stable/modules/ensemble.html (visited on 10/27/2022).

[90] Yandex. *Using the overfitting detector*. URL: https://catboost.ai/en/docs/features/overfitting-detector-desc (visited on 10/28/2022).

A

Data overview

This appendix has the objective of giving an overview of the data. Including the relation between the different features and their distribution.

A.1 Stiffness

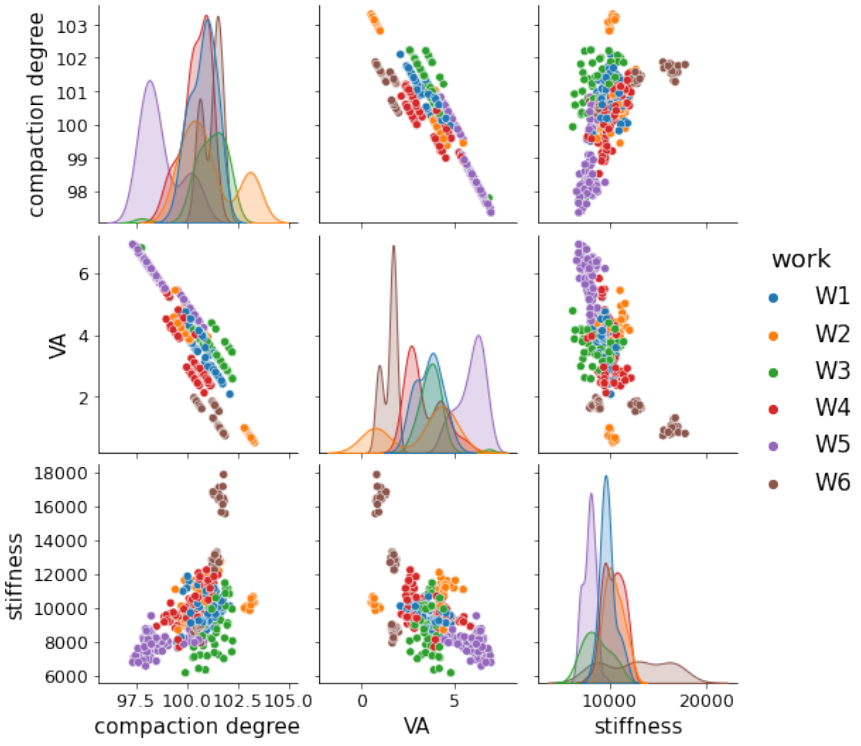


Figure A.1: Composition stiffness overview - 1

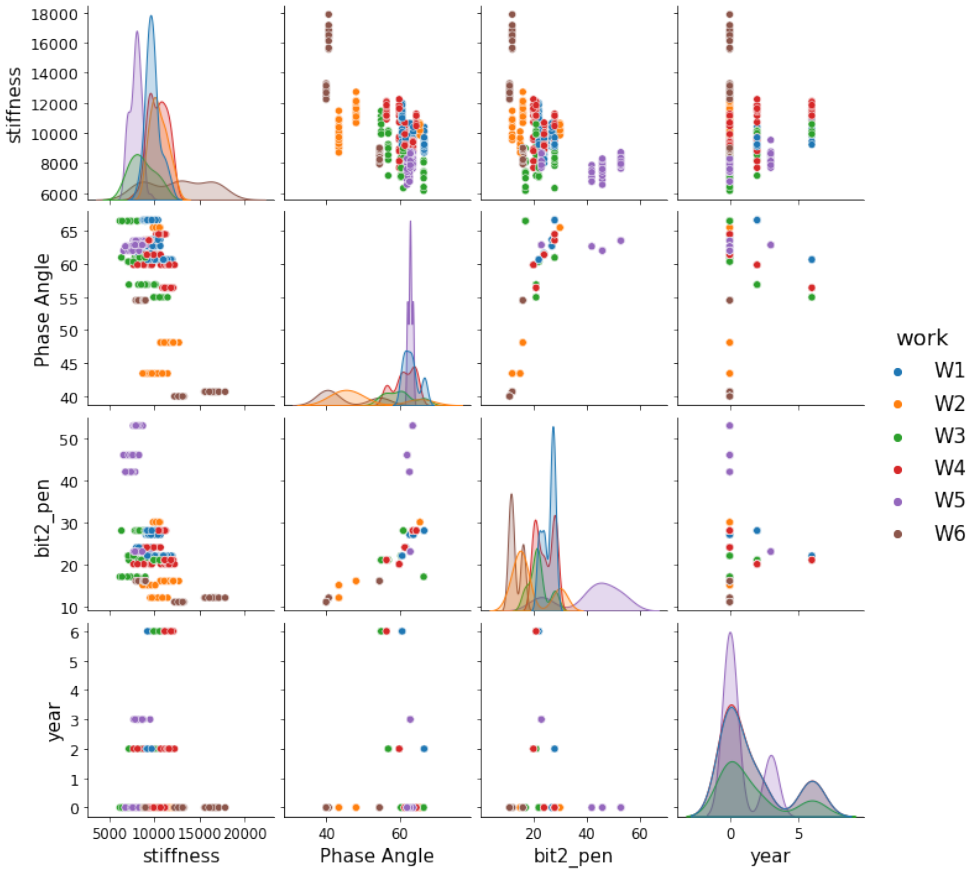


Figure A.2: Composition stiffness overview - 2

A.2 Resistance to fatigue

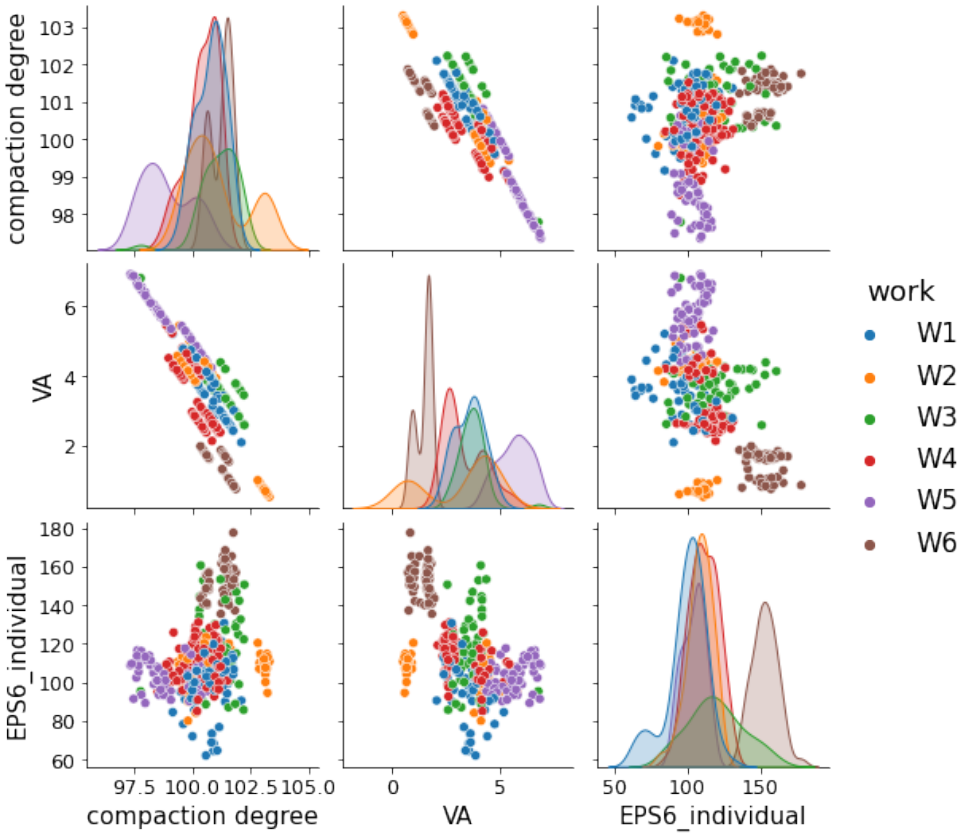


Figure A.3: Composition resistance to fatigue overview - 1

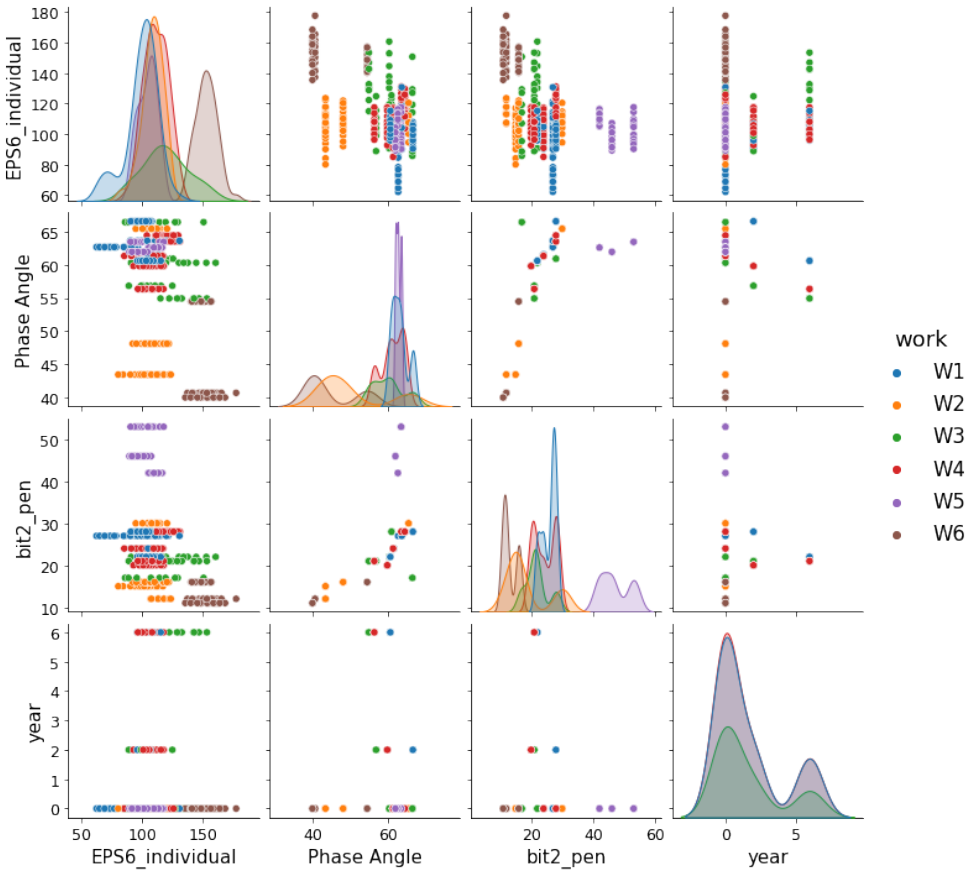


Figure A.4: Composition resistance to fatigue overview - 2

A.3 Resistance to permanent deformation

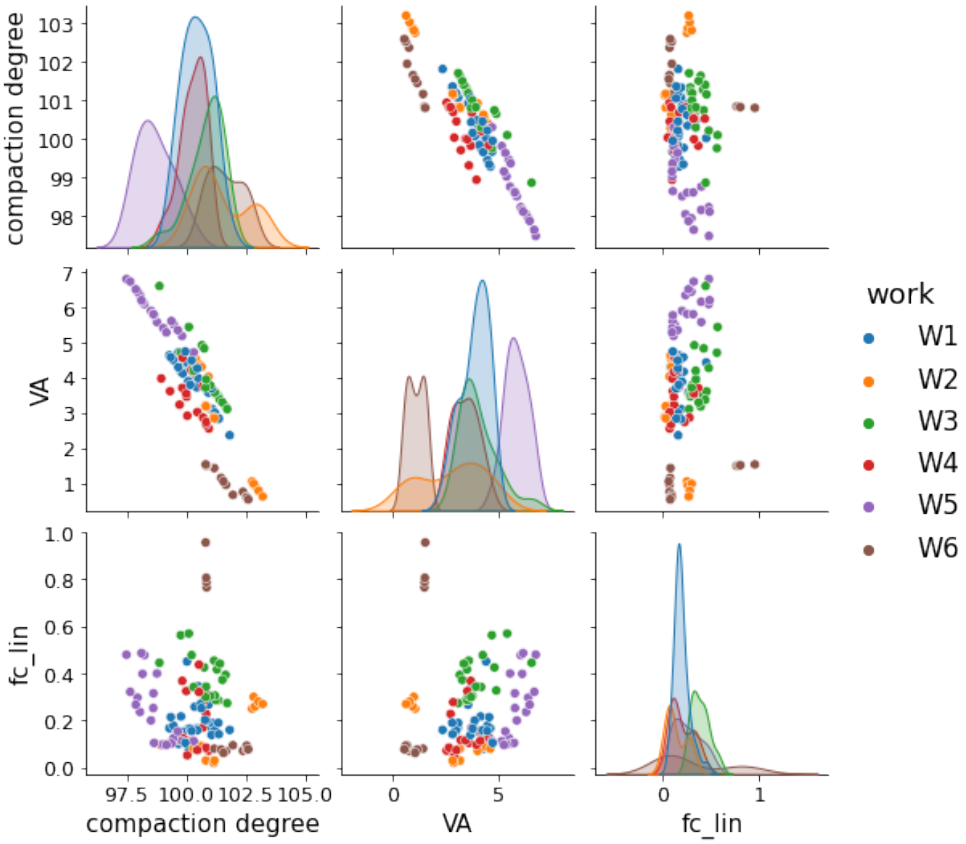


Figure A.5: Composition resistance to permanent deformation overview - 1

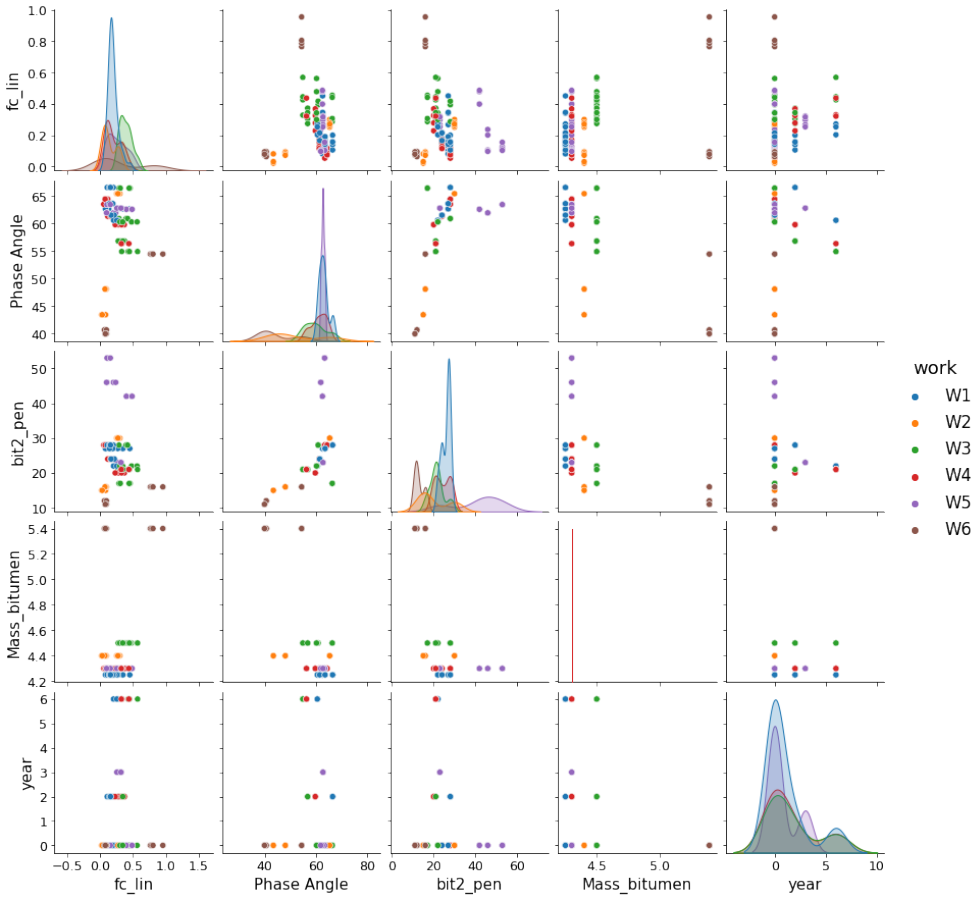


Figure A.6: Composition resistance to permanent deformation overview - 2

A.4 Indirect tensile strength

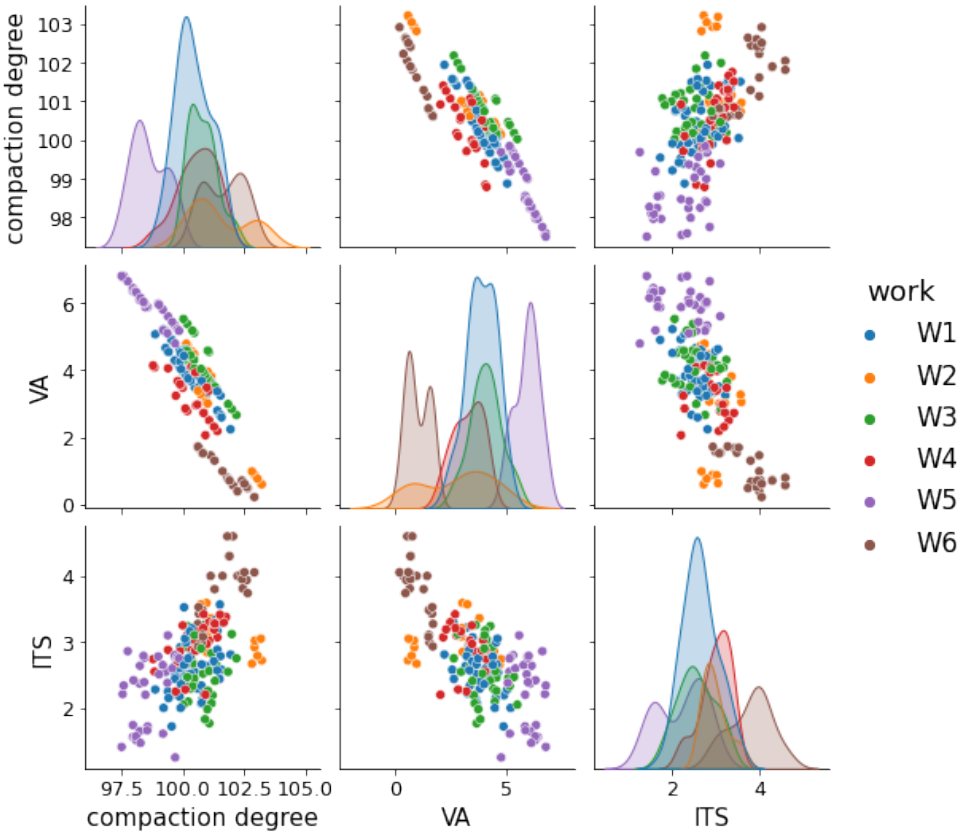


Figure A.7: Composition ITS overview - 1

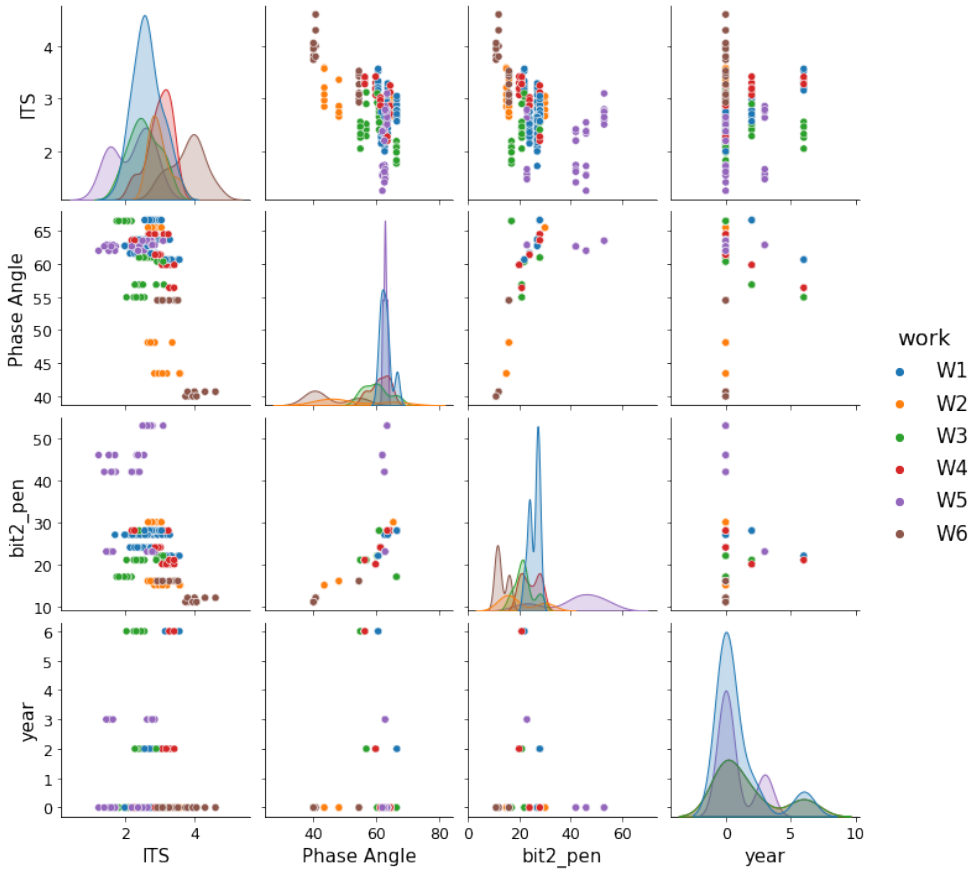


Figure A.8: Composition ITS overview - 2

B

Data preprocessing

This appendix has the objective of showing the tables and plots from the data preprocessing. Including all the availables features, correlation analysis, split and categorical data reduction

The data described in the preprocessing section of Chapter 3 can be found in Table B.1.

Table B.1: Preprocessed raw data

Name	Description
filename	Filename
phase_tot	Abbreviation Work-Phase-Lab-Year
work	Work
phase	Phase
lab	Laboratory
year	Year when measurement took place
monsternames	Monster name
date_prod	Date of production of asphalt
date_prep	Date of specimen preparation
Densities	Density specimen (bulk density)
VA	Air void content
Bitumen 1_TRenK	Temperature Ring and Ball of fresh binder
Bitumen 2_TRenK	Temperature Ring and Ball of recovered binder
Bitumen 1_pen	Penetration of fresh binder
Bitumen 2_pen	Penetration of recovered binder
Bitumen 1_G*	Complex Modulus @ 20řC and 10 rad/s of fresh binder
Bitumen 2_G*	Complex Modulus @ 20řC and 10 rad/s of recovered binder

Bitumen 1_delta	Phase angle @ 20řC and 10 rad/s of fresh binder
Bitumen 2_delta	Phase angle @ 20řC and 10 rad/s of recovered binder
Target density	Target density
RAP percentage	% Reclaimed asphalt percentage in the mix
RAP percentage 1	% Reclaimed asphalt percentage 1 in the mix
RAP percentage 2	% Reclaimed asphalt percentage 2 in the mix
Bitumen percentage	% Bitumen in the mix
Fresh bitumen percentage	% Fresh bitumen in the mix
AG1_percentage_PR	% Bitumen in PR1
AG2_percentage_PR	% Bitumen in PR2
AG1_pen_PR	Penetration of PR1
AG2_pen_PR	Penetration of PR2
AG1_TRenK_PR	Temperature Ring and Ball of PR1
AG2_TRenK_PR	Temperature Ring and Ball of PR2
Target_C22_4	Target composition 22/32
Target_C16	Target composition 16/22
Target_C11_2	Target composition 11/16
Target_C08	Target composition 8/11
Target_C05_6	Target composition 5/8
Target_C002mm	Target composition 2/5
Target_C00063mu	Target composition sand
Target_filler	Target composition filler
Target_bitumen	Target composition bitumen
Extraction_C22_4	Composition after extraction 22/32
Extraction_C16	Composition after extraction 16/22
Extraction_C11_2	Composition after extraction 11/16
Extraction_C08	Composition after extraction 8/11
Extraction_C05_6	Composition after extraction 5/8
Extraction_C002mm	Composition after extraction 2/5
Extraction_C00063mu	Composition after extraction sand
Extraction_filler	Composition after extraction filler
Extraction_bitumen	Composition after extraction bitumen
Reference_C22_4	Reference composition 22/32
Reference_C16	Reference composition 16/22
Reference_C11_2	Reference composition 11/16
Reference_C08	Reference composition 8/11
Reference_C05_6	Reference composition 5/8

Reference_C002mm	Reference composition 2/5
Reference_C00063mu	Reference composition sand
Reference_filler	Reference composition filler
Reference_bitumen	Reference composition bitumen
dry_wet	Dry of wet subset
type_setup	Type of setup
descr_setup	Descripton of setup
mix_setup	Type of mixer
comp_setup	Type of compactor
frict_red_system	Friction reduction system
Stiffness	Stiffness
ε	Strain level
S_{50}	Half of initial stiffness
N_{50}	Number of cycles at half stiffness
ε_6	Epsilon 6
$\varepsilon_6(Individual)$	Individual calculated Epsilon6 value
A_1 lin	A1 - constant linear fit
B_1 lin	B1 - constant linear fit
R_2 lin	Regression coefficient linear fit
f_c lin	fc (creep rate) $\frac{\mu\varepsilon}{cycles}$
A_{log}	A - constant logarithmic fit
B_{log}	B - constant logarithmic fit
$R2_{log}$	Regression coefficient logarithmic fit
ε_{1000} log	Permanent deformation after 1000 loading cycles
ITS	Indirect tensile strength @ 15řC

Table B.2: Discarded Features

Name	**Reason**	**Name**	**Reason**
filename	Descriptive	Extraction_C00063mu	- Missing values - Tests needed
phase_tot	Descriptive	Extraction_filler	- Missing values - Tests needed
Lab	Descriptive	Extraction_bitumen	- Missing values - Tests needed
monsternames	Descriptive	Reference_C22_4	Missing values
date_prod	Descriptive	Reference_C16	Missing values
date_prep	Descriptive	Reference_C11_2	Missing values

bit1_TRenK	Incorporated	Reference_C08	Missing values
bit1_pen	Incorporated	Reference_C05_6	Missing values
bit1_delta	Incorporated	Reference_C002mm	Missing values
percentage_PR_1	Incorporated	Reference_C00063mu	Missing values
percentage_PR_2	Incorporated	Reference_filler	Missing values
percentage_bit	Tests needed	Reference_bitumen	Missing values
percentage_fresh_bit	Incorporated	type_setup	Descriptive
AG1_percentage_PR	Incorporated	descr_setup	Descriptive
AG2_percentage_PR	Incorporated	eps	Tests needed
AG1_pen_PR	Incorporated	S50	Tests needed
AG2_pen_PR	Incorporated	N50	Tests needed
AG1_TRenK_PR	Incorporated	EPS6	Tests needed
AG2_TRenK_PR	Incorporated	A1_lin	Tests needed
Extraction_C22_4	- Missing values - Tests needed	B1_lin	Tests needed
Extraction_C16	- Missing values - Tests needed	R2_lin	Tests needed
Extraction_C11_2	- Missing values - Tests needed	A_log	Tests needed
Extraction_C08	- Missing values - Tests needed	B_log	Tests needed
Extraction_C05_6	- Missing values - Tests needed	R2_log	Tests needed
Extraction_C002mm	- Missing values - Tests needed	eps1000_log	Tests needed

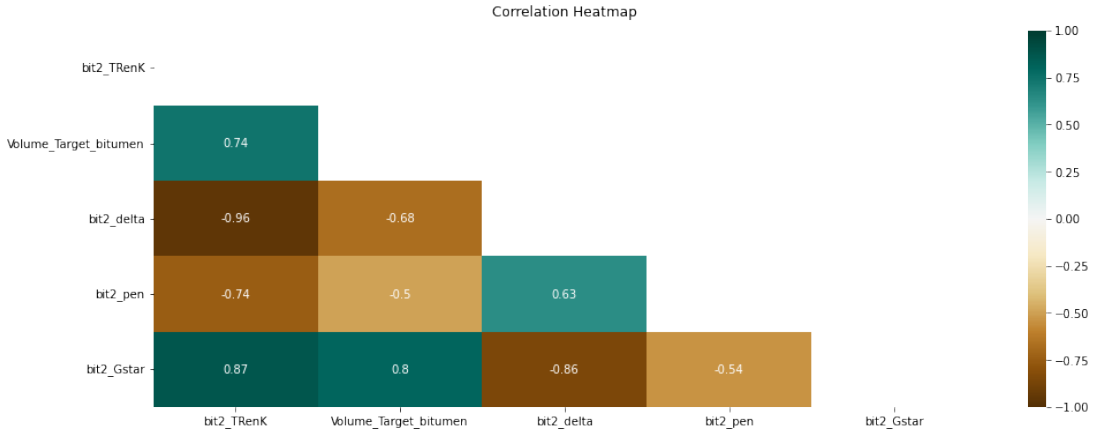


Figure B.1: Bitumen correlation

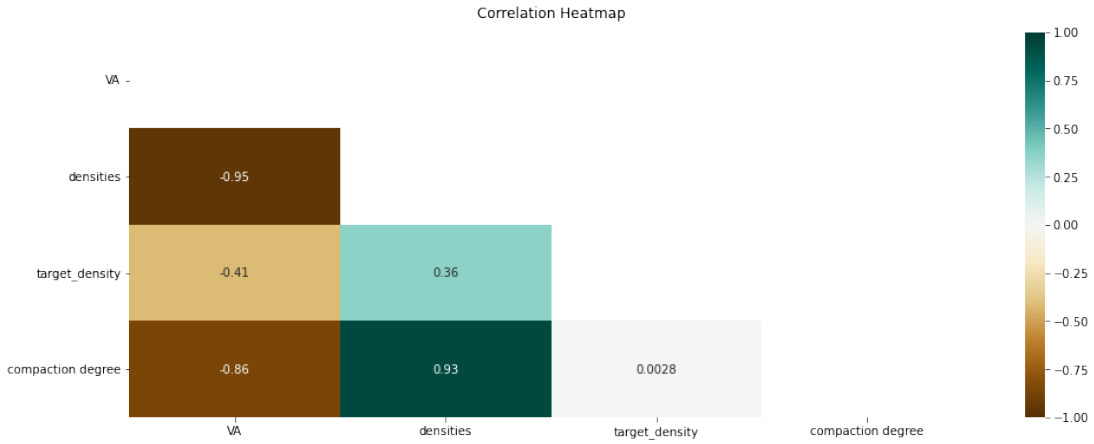


Figure B.2: Density correlation

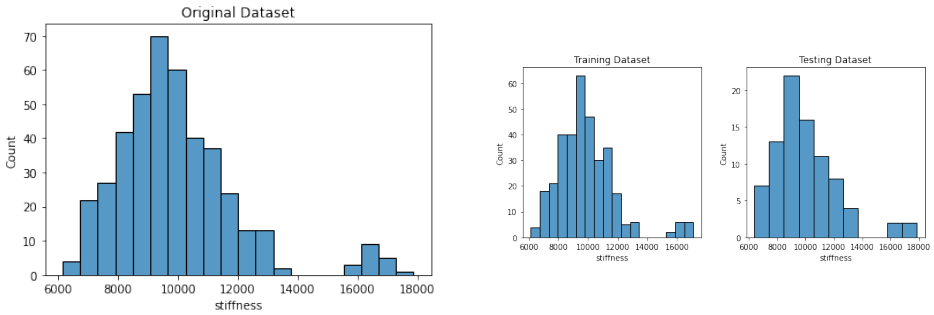


Figure B.3: Stiffness distribution

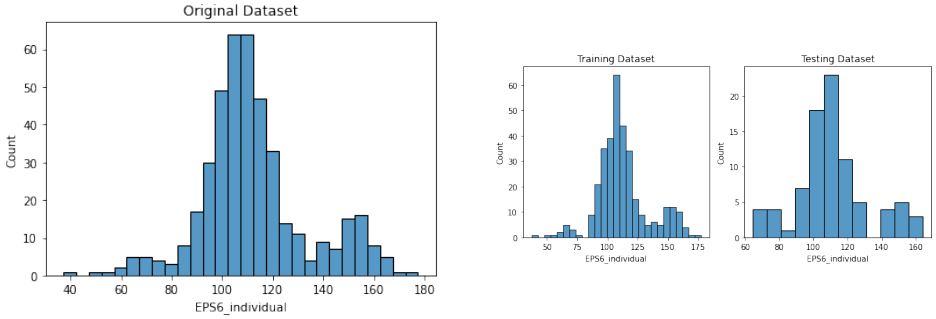


Figure B.4: ε_6 distribution

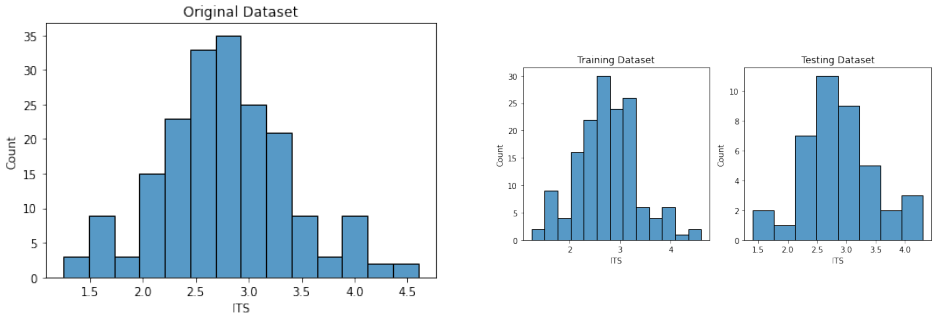


Figure B.5: ITS distribution

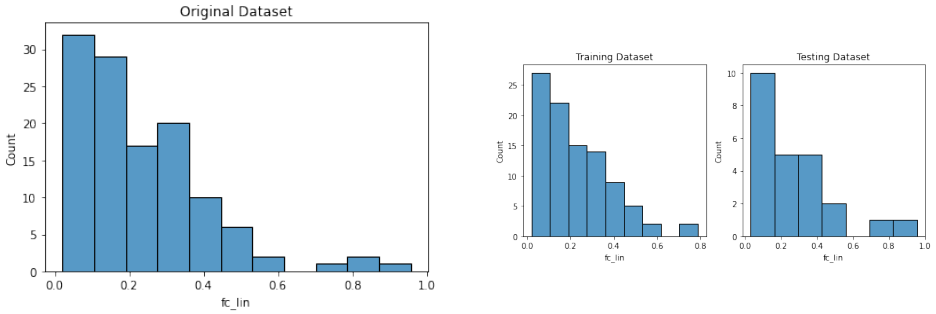


Figure B.6: f_c lin distribution

Table B.3: Compaction set-up data

Original name	Coherent group name
Field roller	Field roller
Walsverdichting (HAMM drierol)	
Gyrator	Gyrator compactor
Gyrator compactor	
Hand roller	Hand roller
Plaatverdichter	Segment compactor
Segment compactor	
Shear box	Shear box

Table B.4: Mixing set-up data

Original name	Coherent group name
Asphalt plant	Asphalt plant
Molen gemengd	
molengemengd	
Wibau W.S.T. 150 charge menger	Forced action mixer
dwangmenger Lab	
Forced action mixer	Planetory mixer
Hobart planeetmenger	

Table B.5: Friction reduction system data

Original name	Coherent group name
2 maal Marshallpapier	2 maal Marshallpapier
2 rubbermembranen met siliconenvet per zijde	Two layer rubber with silicon grease
latex en silicoon film	
Latexmembranen met siliconenvet	
two layer rubber	
PTFE gecoat vlies	PTFE gecoat vlies
Acre system	Acre system

C

SHAP importances

The SHAP importances are tables with the SHAP value from the feature importance located in the results chapter.

C.1 Stiffness and resistance to fatigue

Table C.1: Stiffness and resistance to fatigue feature importance

Features	Mean SHAP Value
Compaction Degree	348.977642
Bitumen Phase Angle	335.374508
Mixing Set-up	188.059042
Bitumen Penetration	183.210488
Target Sand Mass Percentage	149.857387
Target Density	145.854286
Target Stone Mass Percentage	141.713306
Year	133.742498
RAP Percentage	100.641101
Target Filler Mass Percentage	92.695177
Target Bitumen Mass Percentage	64.116634
Compaction Set-up	52.486127

Features	Mean SHAP Value
Target Bitumen Mass Percentage	2.888558
Target Filler Mass Percentage	2.328424
Target Density	2.129955
Compaction Set-up	1.717966
Compaction Degree	1.338033
Bitumen Phase Angle	1.12459
RAP Percentage	1.069403
Mixing Set-up	0.921308
Target Stone Mass Percentage	0.76657
Target Sand Mass Percentage	0.7636
Bitumen Penetration	0.565547
Year	0.301376

C.2 Resistance to permanent deformation and ITS

Table C.2: Resistance to permanent deformation and ITS feature importance

Features	Mean SHAP Value
Compaction Set-up	0.026114
Friction Reduction System	0.02114
Target Bitumen Mass Percentage	0.015579
Target Filler Mass Percentage	0.014584
Mixing Set-up	0.011891
Bitumen Phase Angle	0.009859
Target Density	0.007667
Bitumen Penetration	0.007204
Target Sand Mass Percentage	0.005161
Compaction Degree	0.005105
RAP Percentage	0.004332
Year	0.003837
Target Stone Mass Percentage	0.003289

Features	Mean SHAP Value
Sample Condition	0.108333
Bitumen Phase Angle	0.102627
Compaction Degree	0.08196
Target Density	0.078112
Bitumen Penetration	0.038085
RAP Percentage	0.036806
Mixing Set-up	0.030177
Target Bitumen Mass Percentage	0.026631
Target Filler Mass Percentage	0.02582
Compaction Set-up	0.023145
Target Stone Mass Percentage	0.022866
Target Sand Mass Percentage	0.022628
Year	0.01314

D

Sensitivity Analysis per work

This appendix has the objective to give an overview of the SHAP summary plots for each of the works available in the dataset

D.1 Work 1

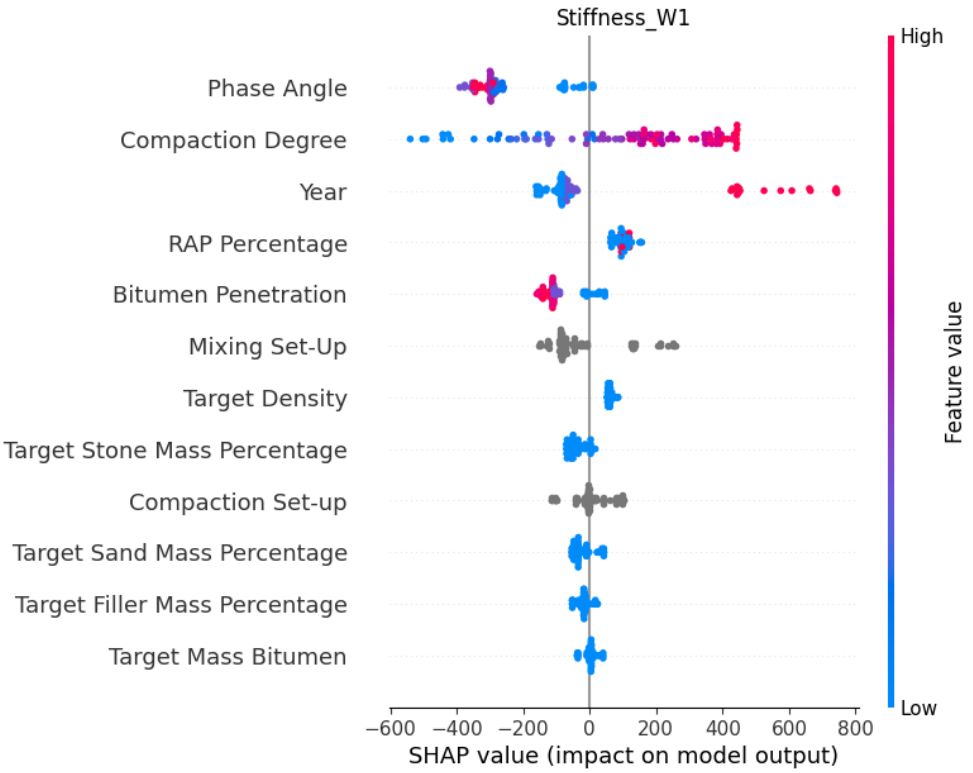


Figure D.1: Work 1 stiffness summary plot

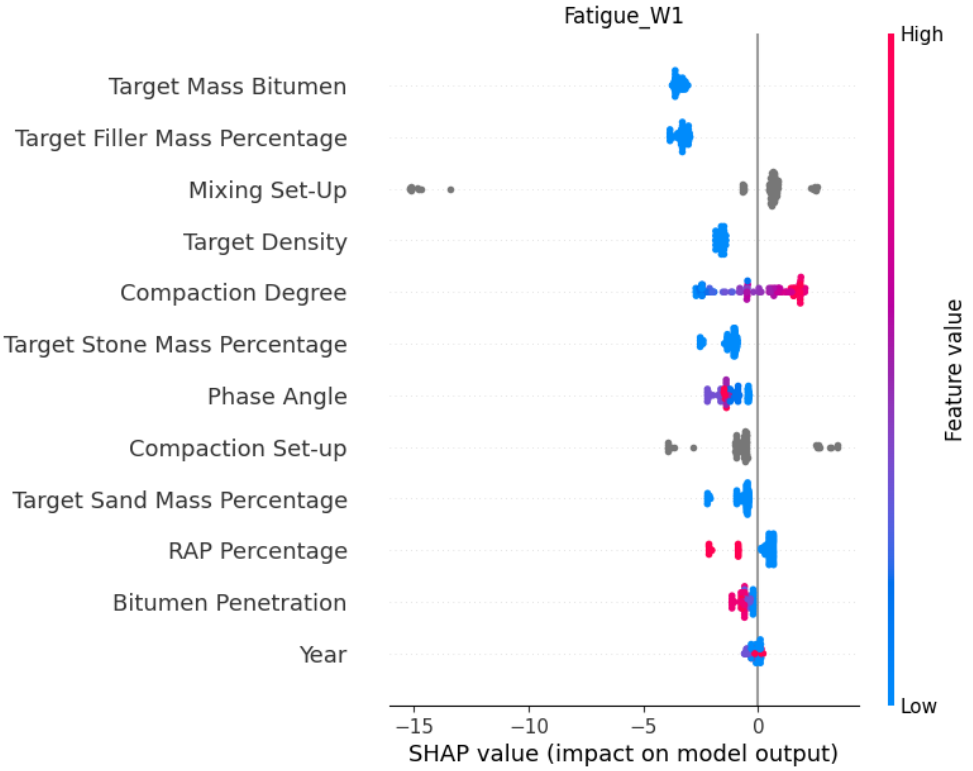


Figure D.2: Work 1 fatigue summary plot

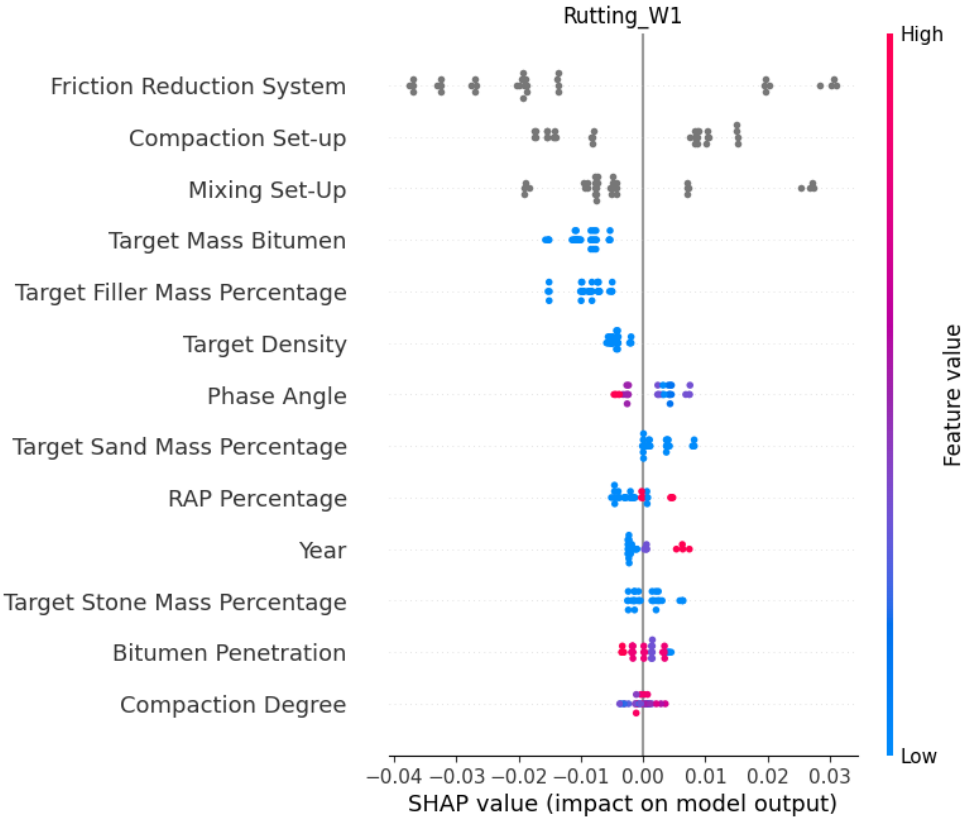


Figure D.3: Work 1 rutting summary plot

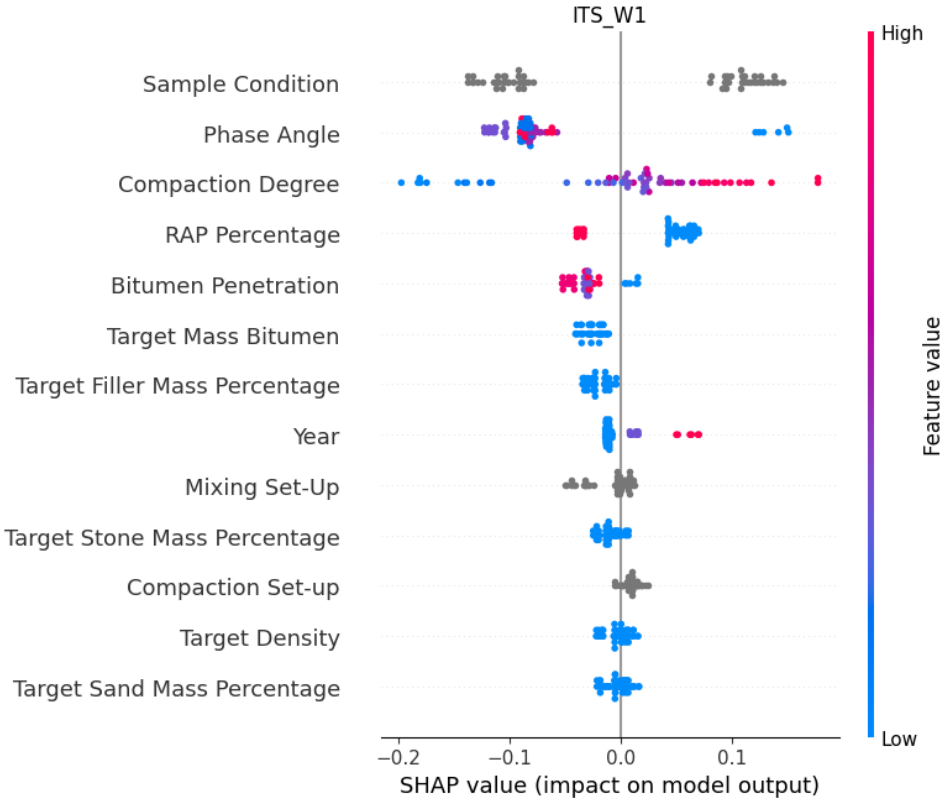


Figure D.4: Work 1 ITS summary plot

D.2 Work 2

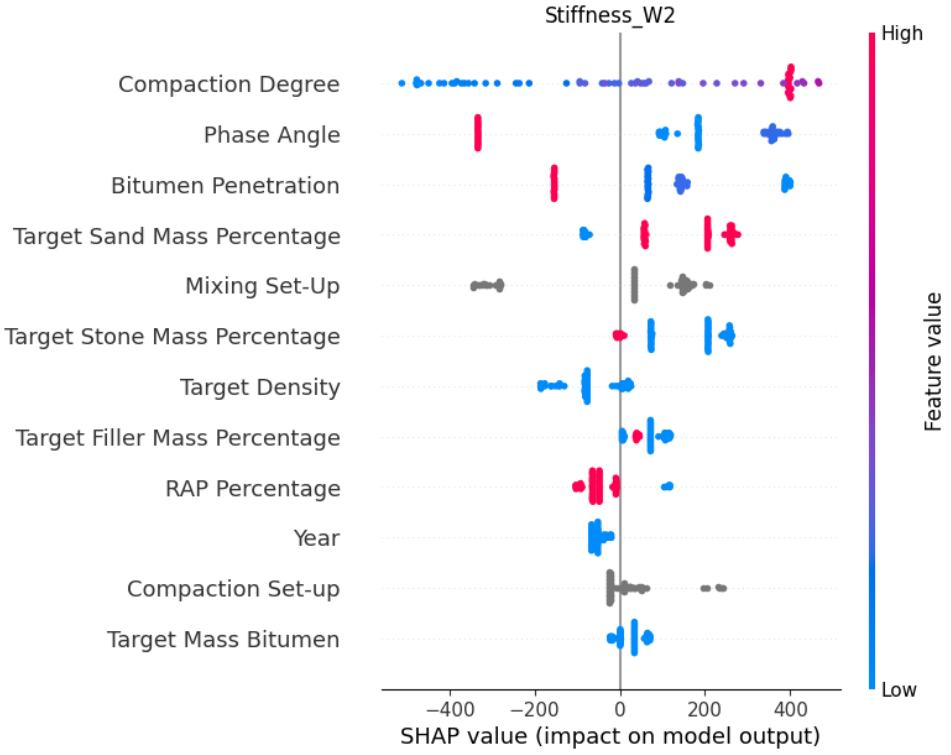


Figure D.5: Work 2 stiffness summary plot

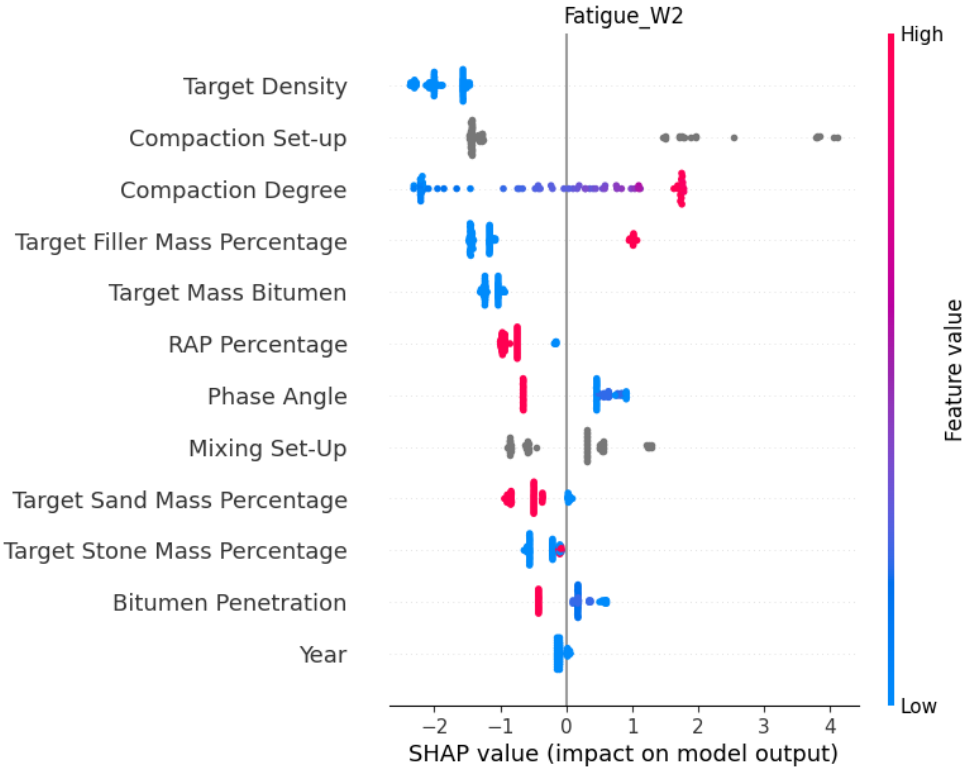


Figure D.6: Work 2 fatigue summary plot

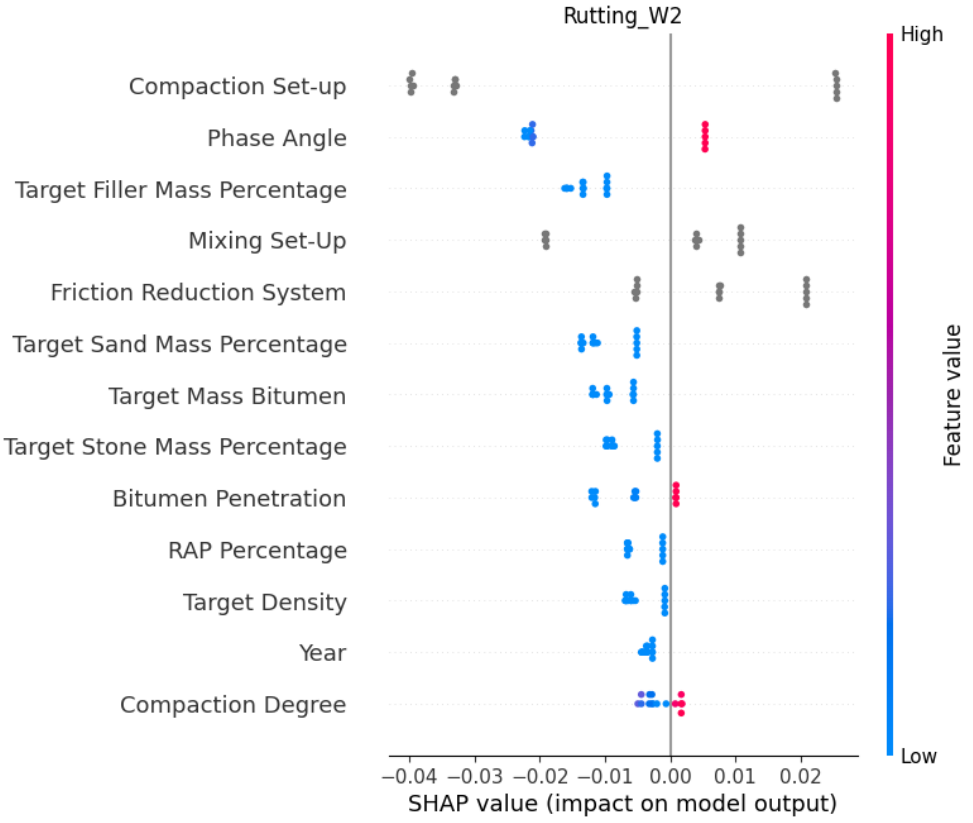


Figure D.7: Work 2 rutting summary plot

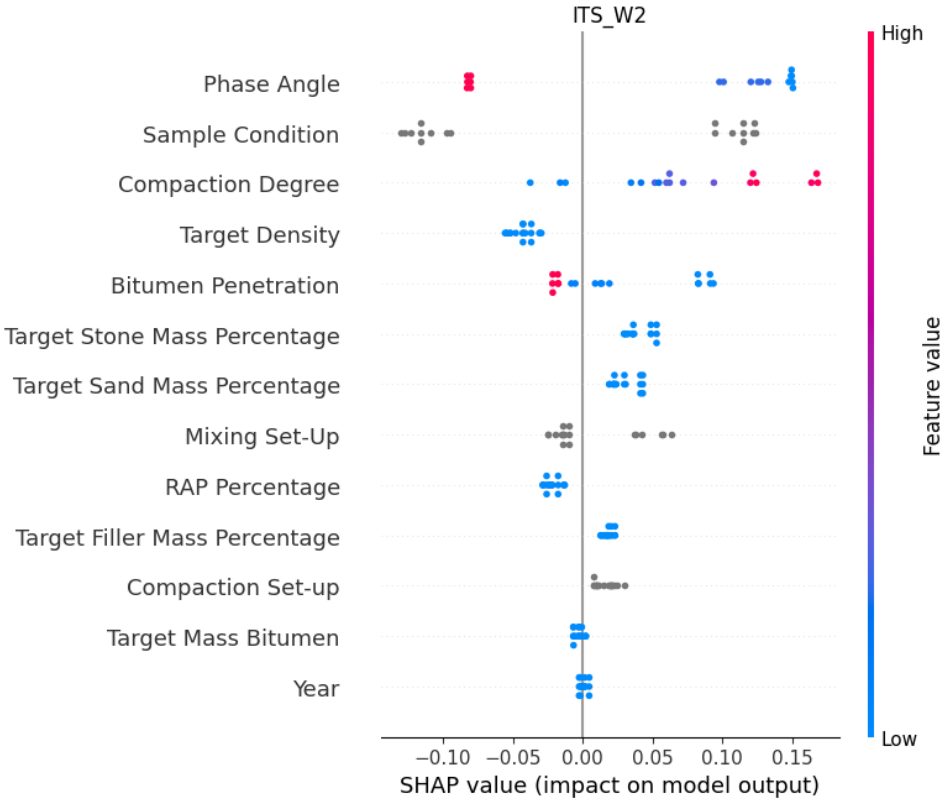


Figure D.8: Work 2 ITS summary plot

D.3 Work 3

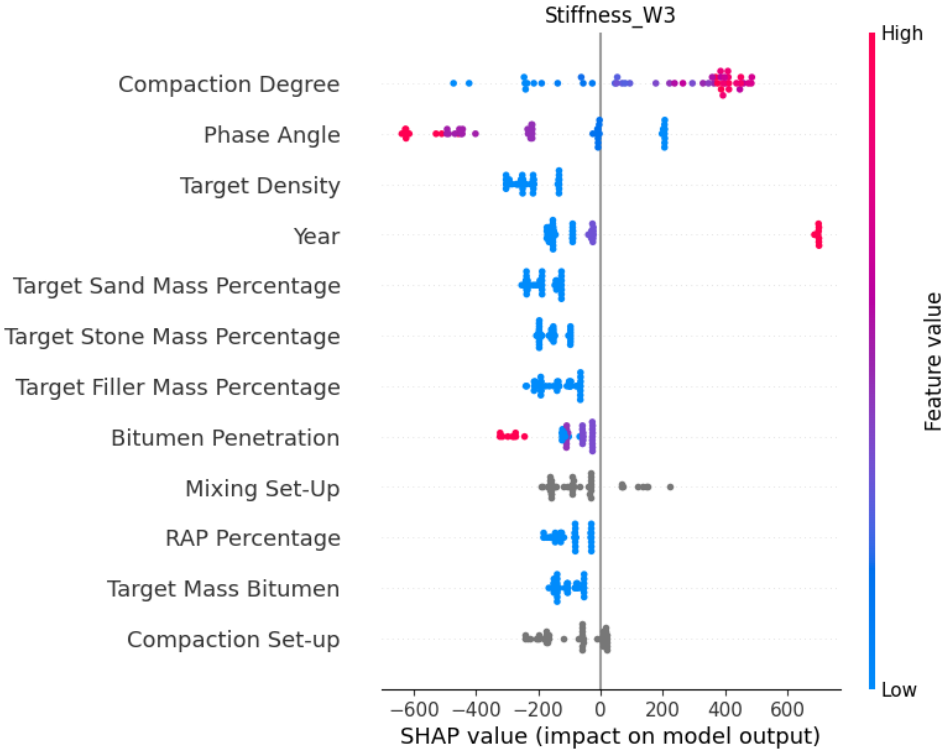


Figure D.9: Work 3 stiffness summary plot

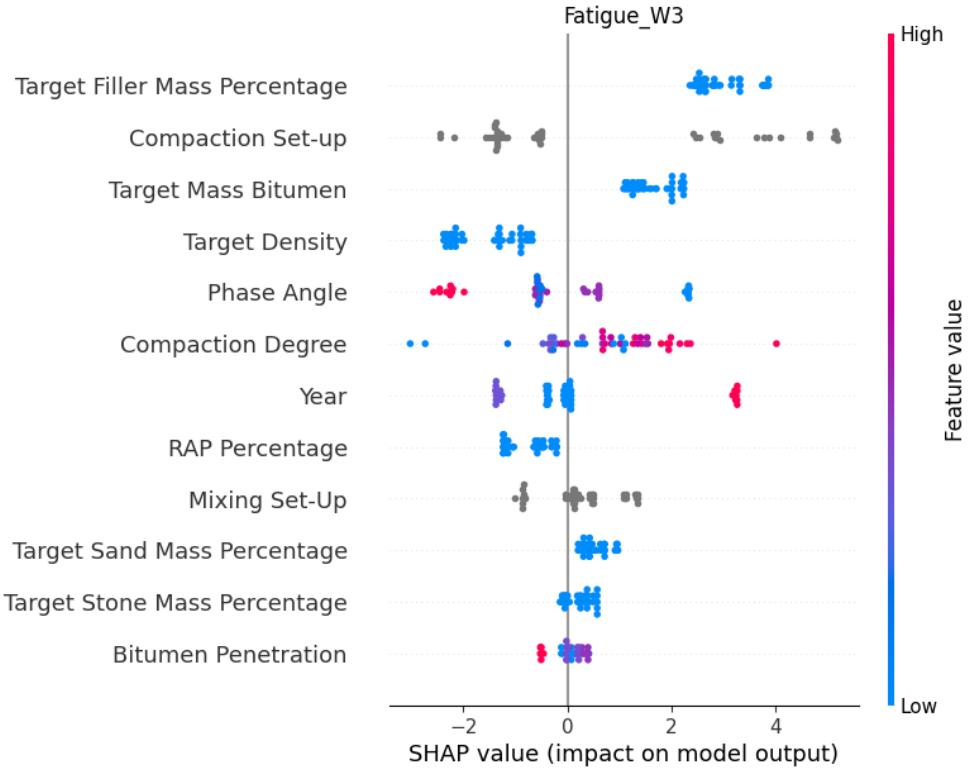


Figure D.10: Work 3 fatigue summary plot

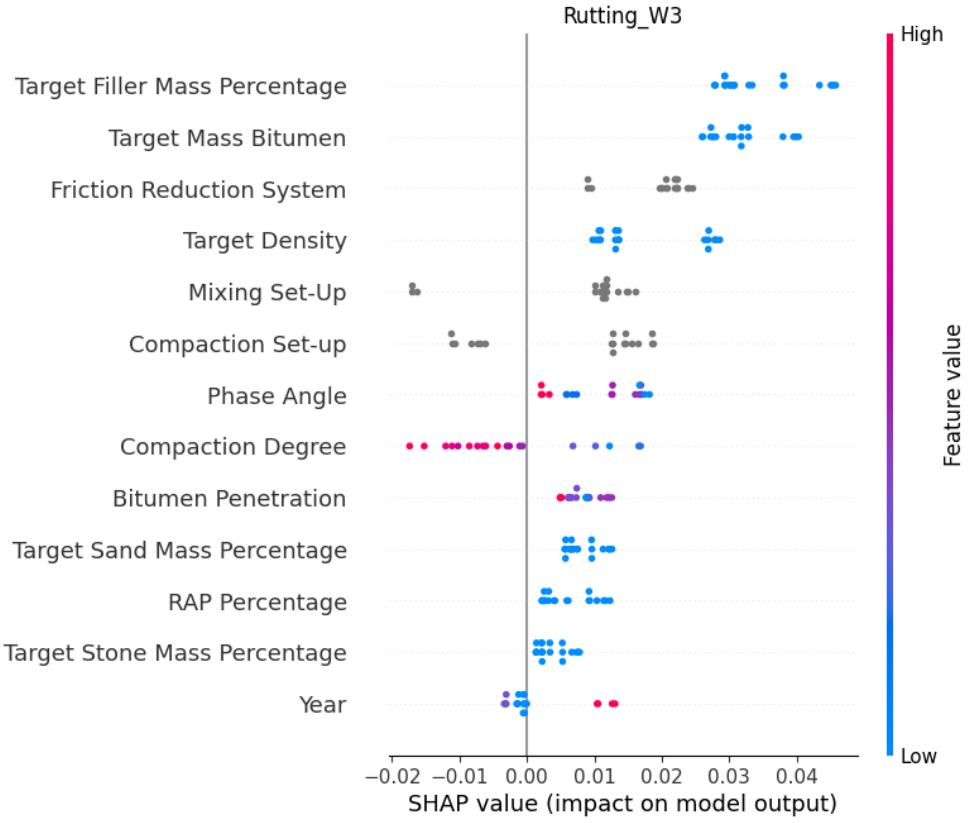


Figure D.11: Work 3 rutting summary plot

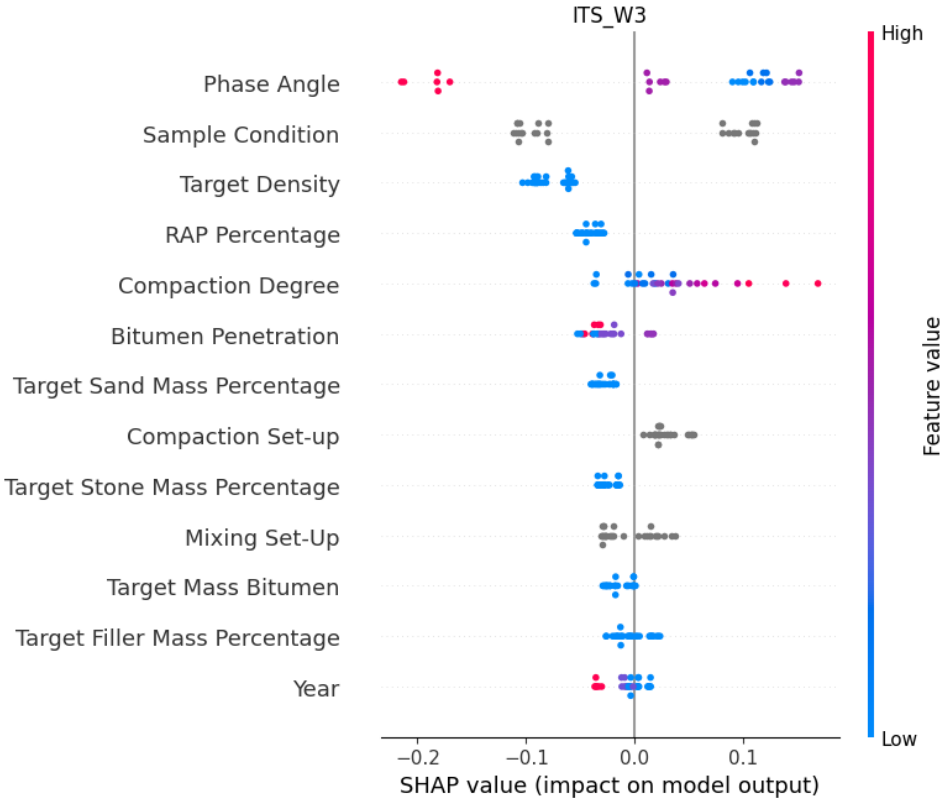


Figure D.12: Work 3 ITS summary plot

D.4 Work 4

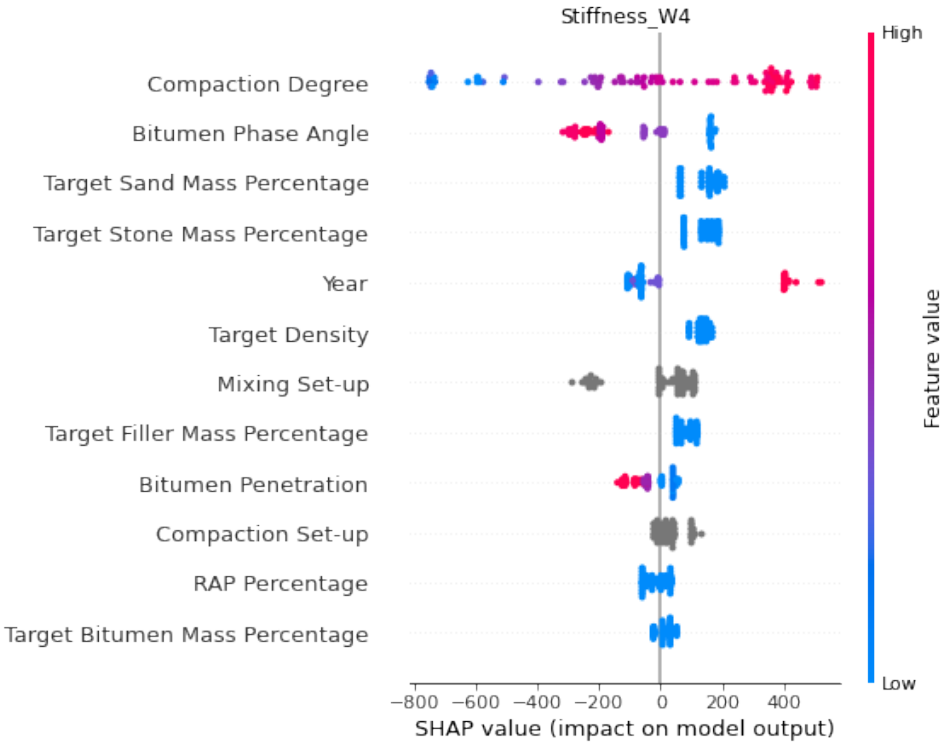


Figure D.13: Work 4 stiffness summary plot

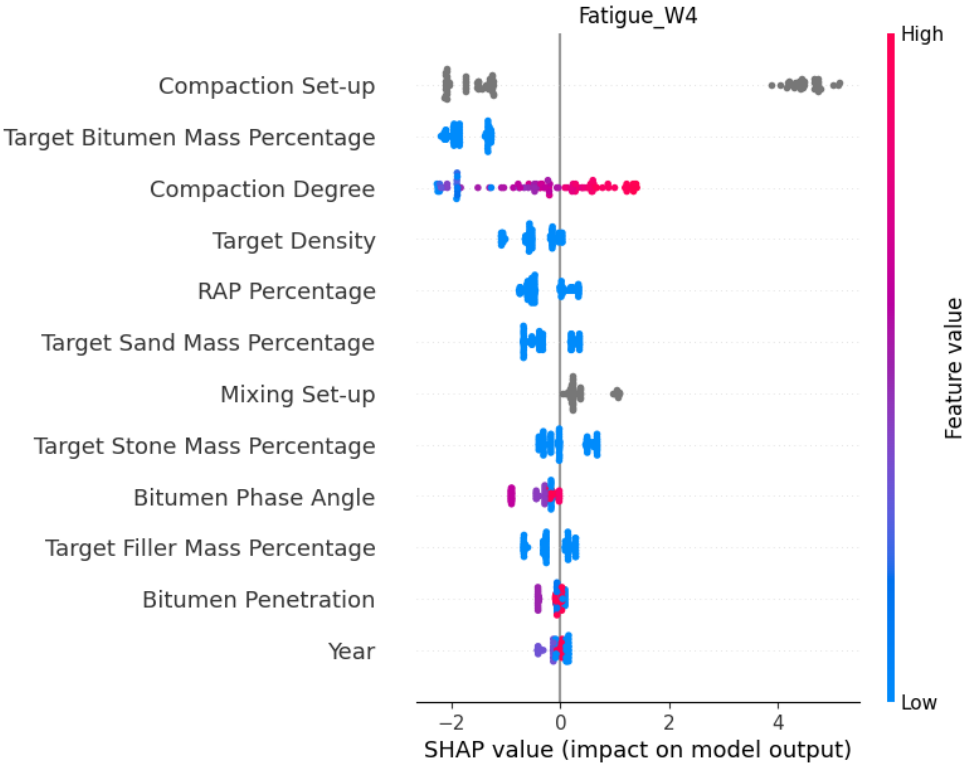


Figure D.14: Work 4 fatigue summary plot

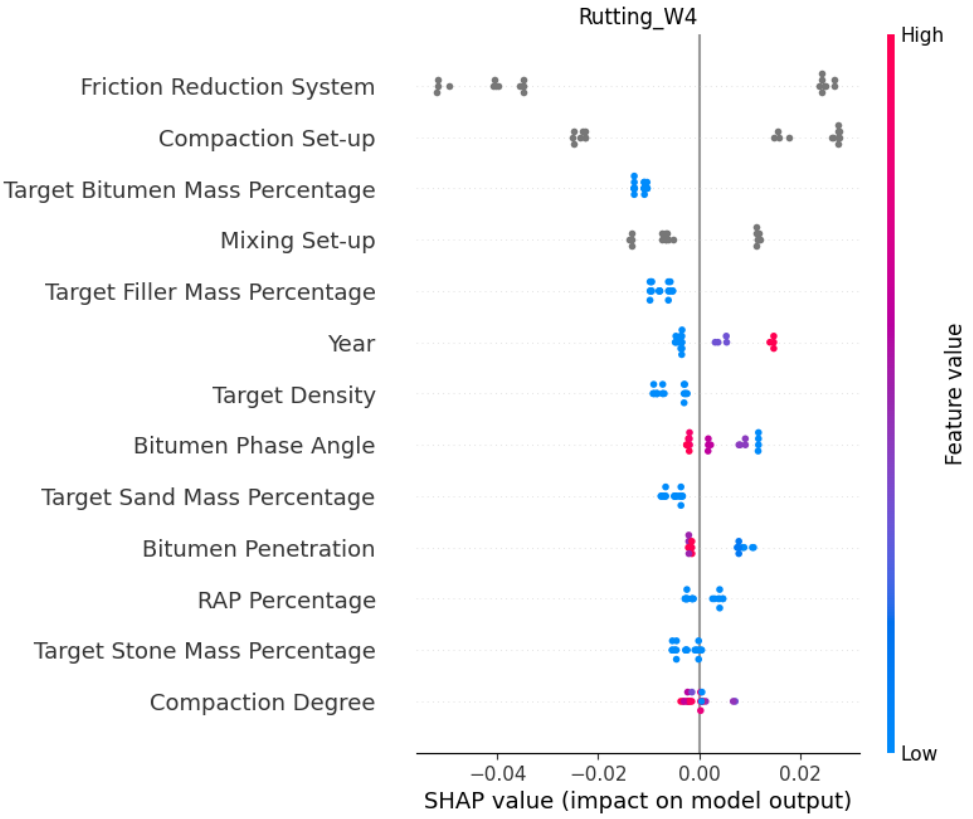


Figure D.15: Work 4 rutting summary plot

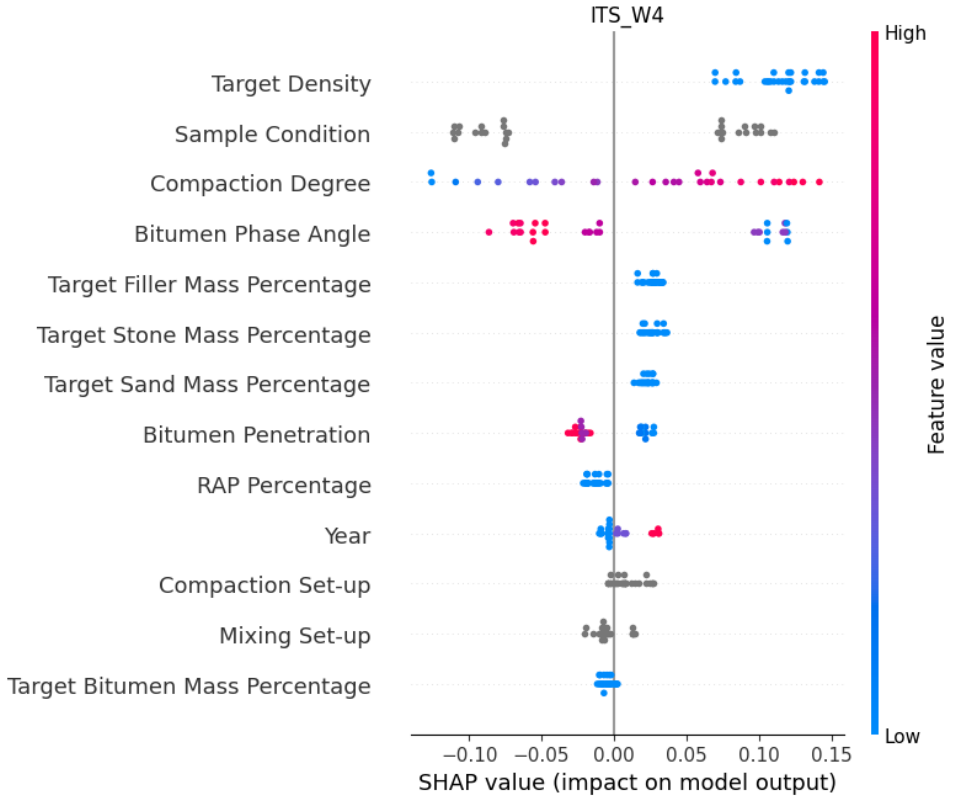


Figure D.16: Work 4 ITS summary plot

D.5 Work 5

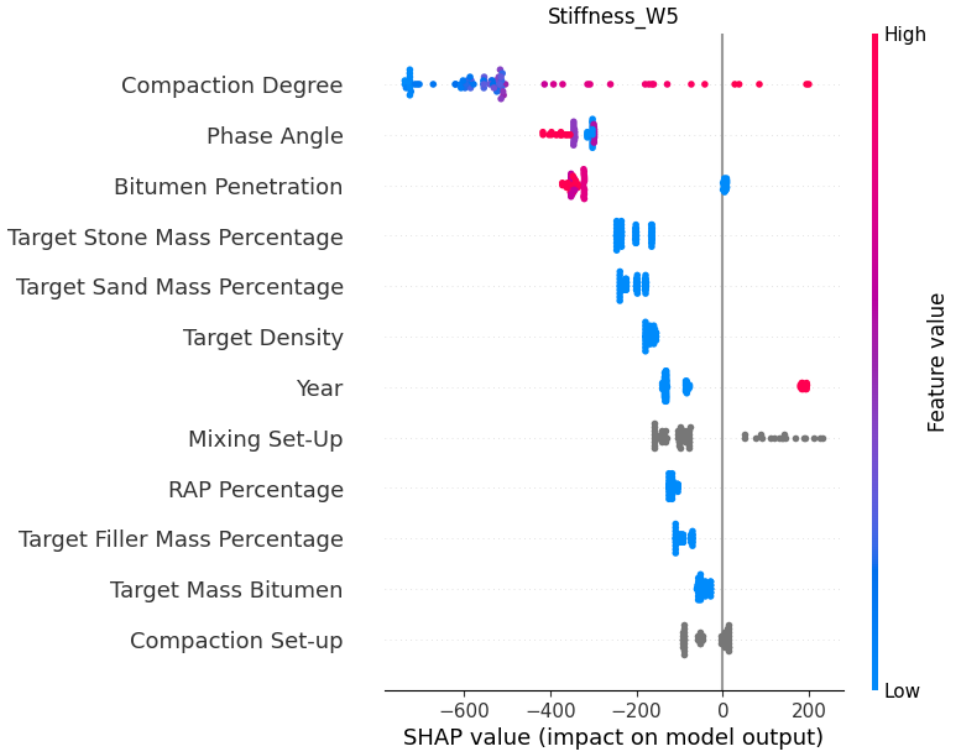


Figure D.17: Work 5 stiffness summary plot

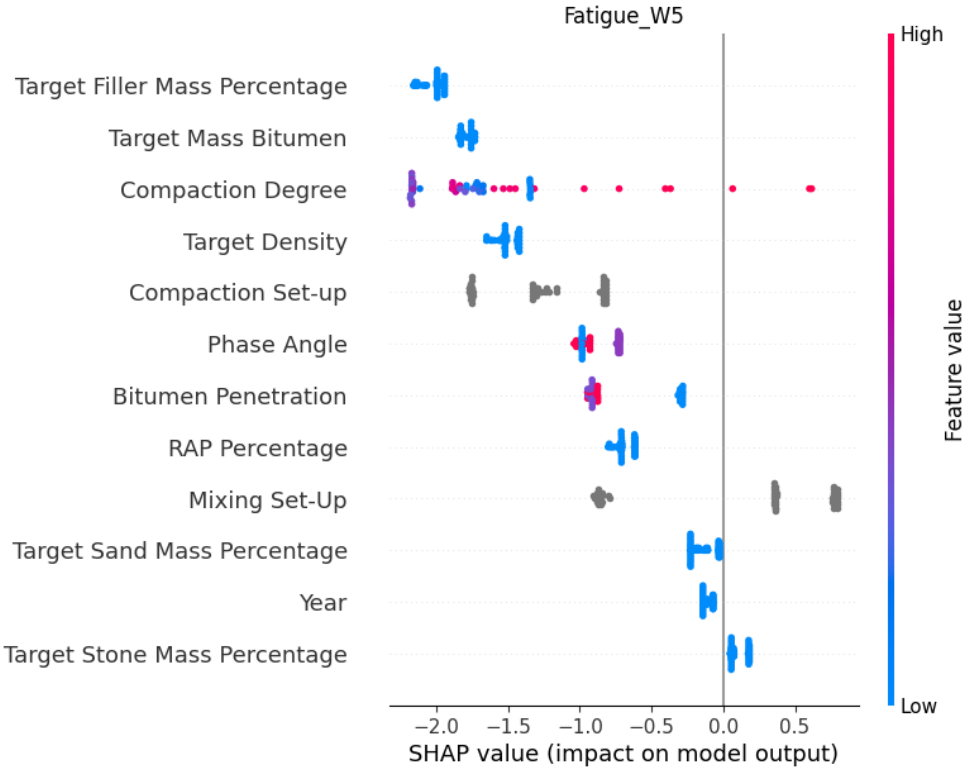


Figure D.18: Work 5 fatigue summary plot

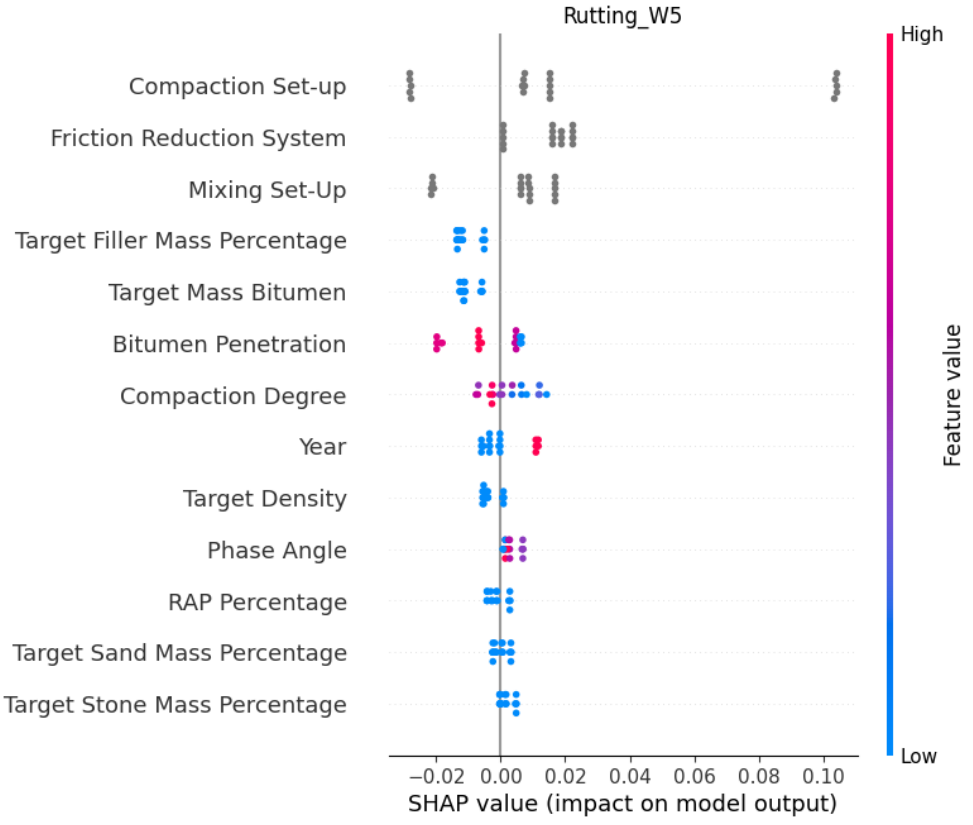


Figure D.19: Work 5 rutting summary plot

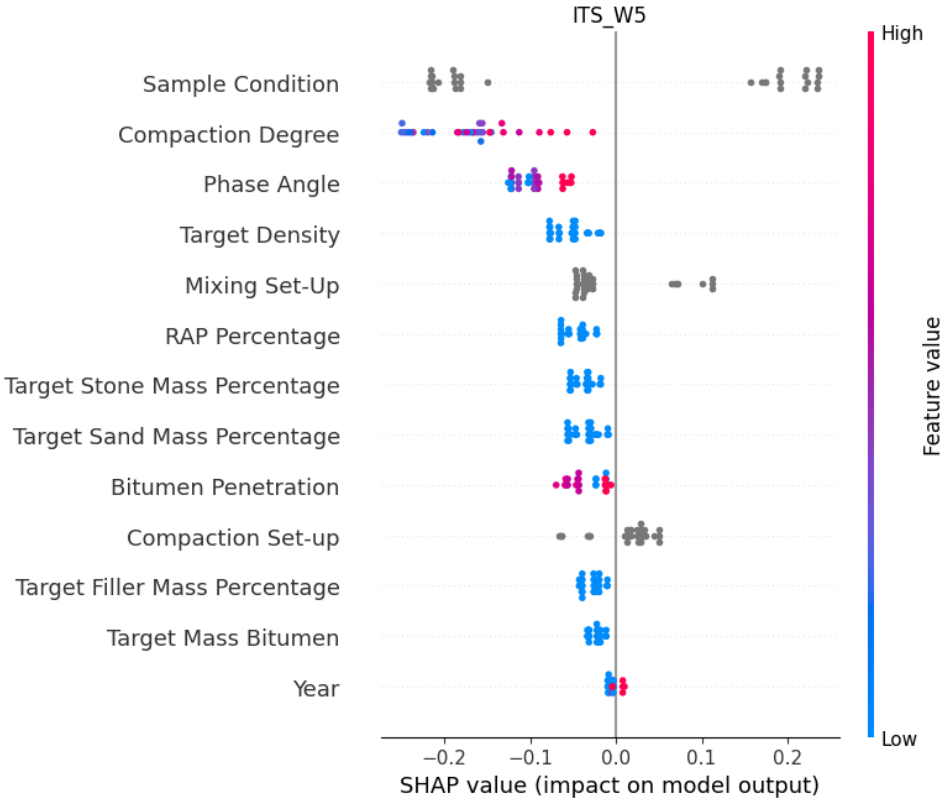


Figure D.20: Work 5 ITS summary plot

D.6 Work 6

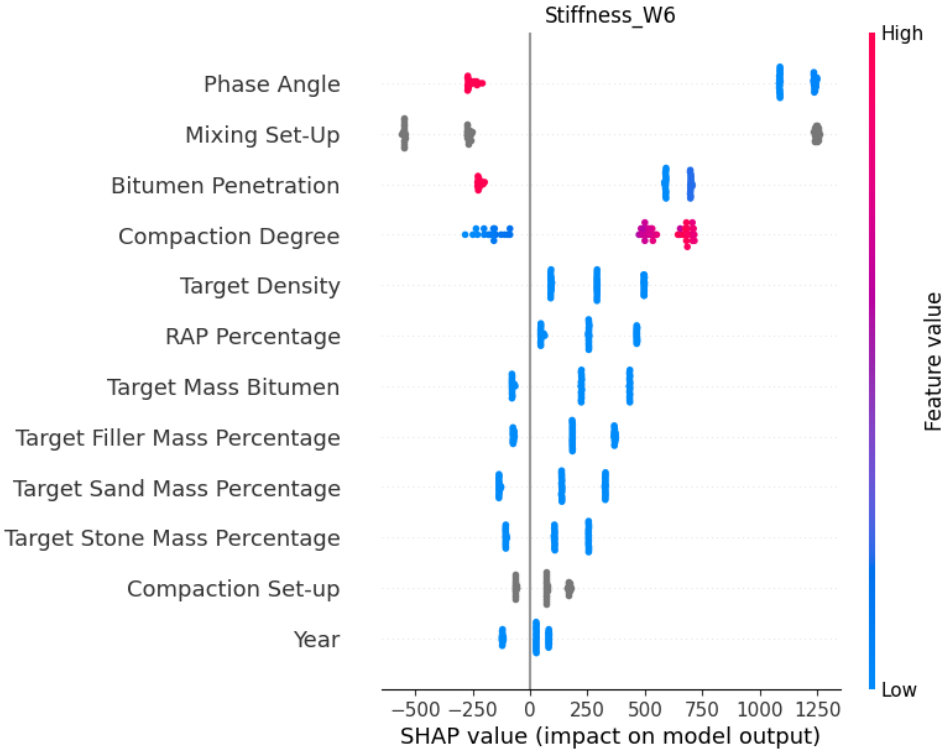


Figure D.21: Work 6 stiffness summary plot

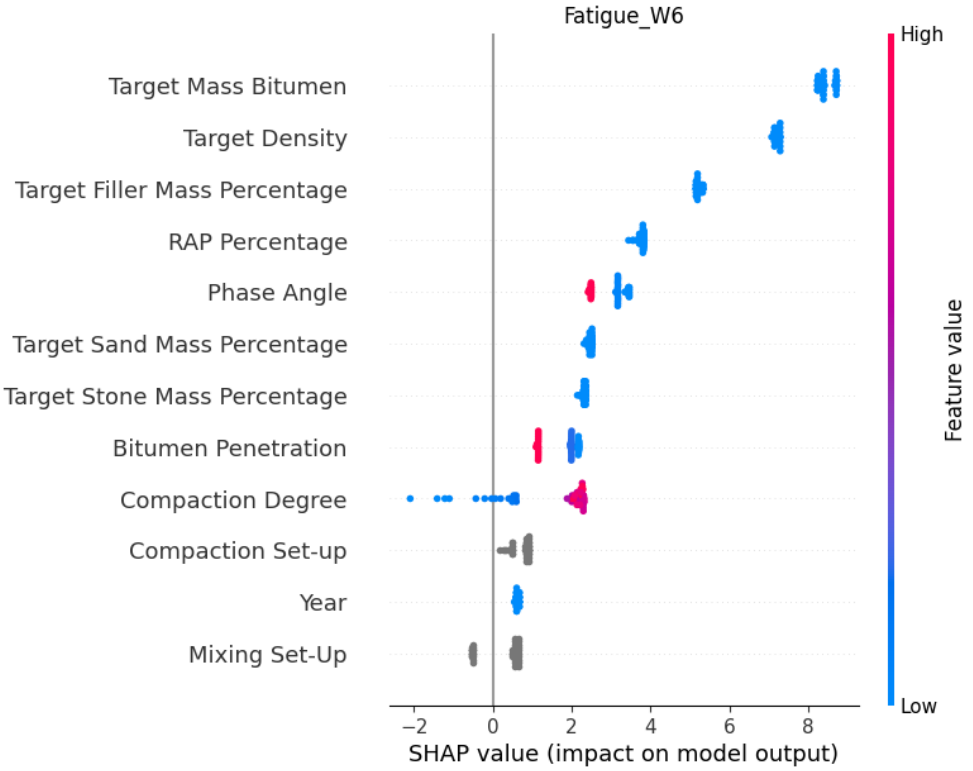


Figure D.22: Work 6 fatigue summary plot

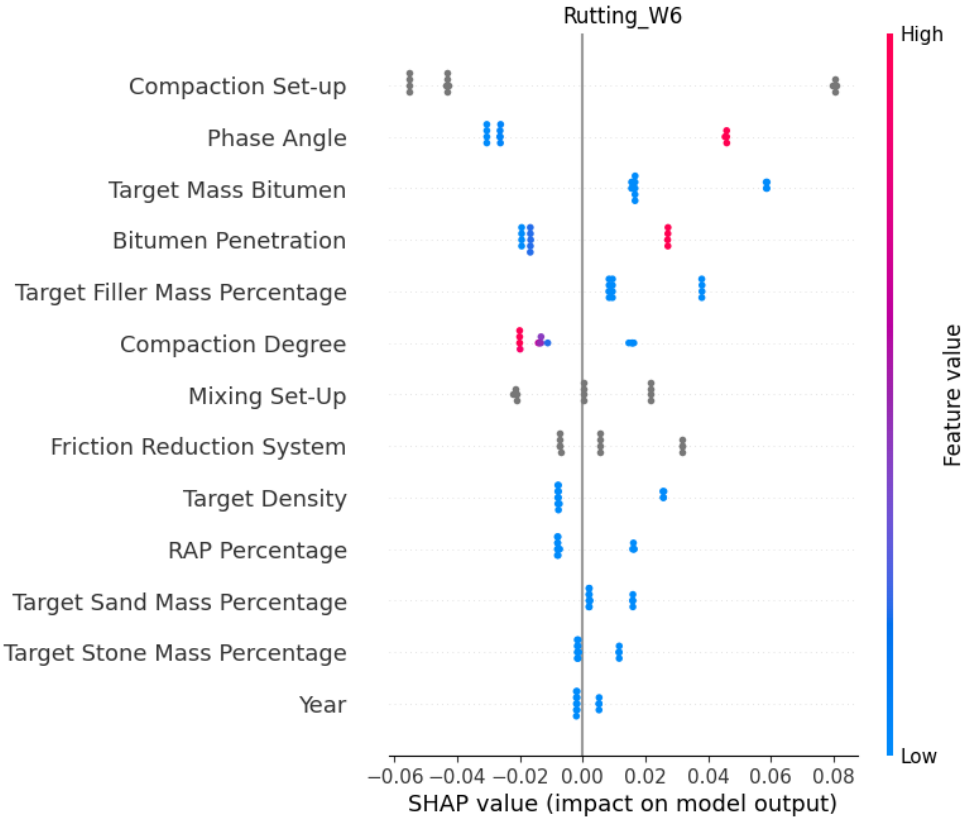


Figure D.23: Work 6 rutting summary plot

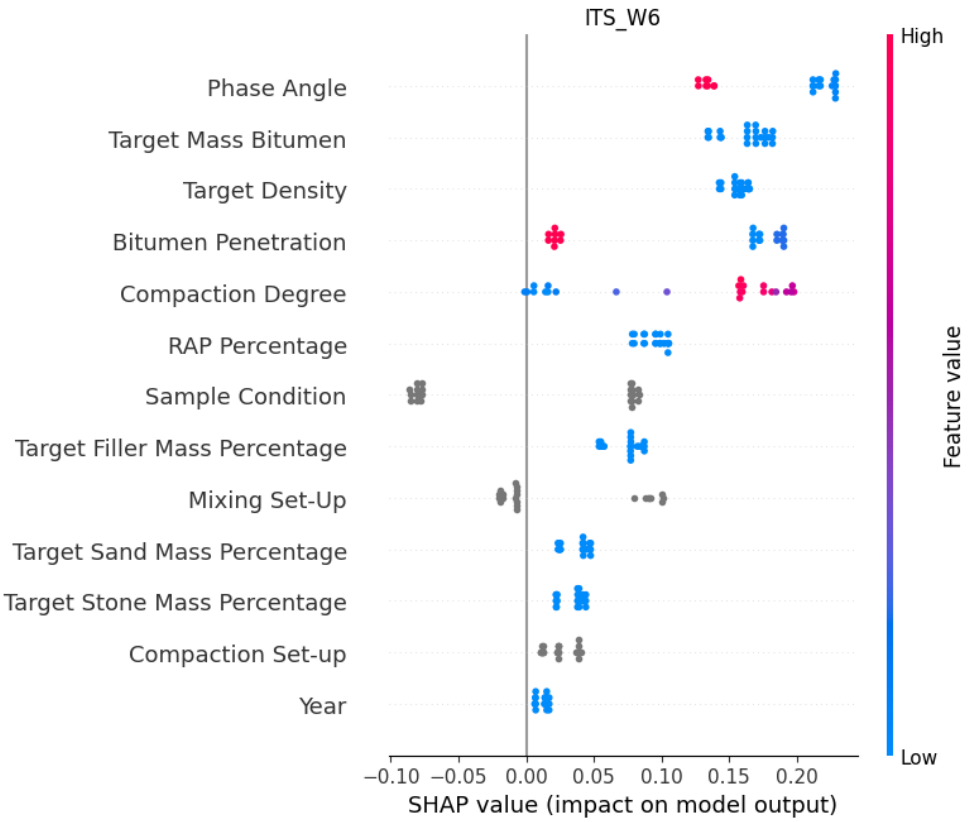


Figure D.24: Work 6 ITS summary plot

E

Algorithm Examples

E.1 Code Example

Importing the libraries

[ ]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import os
import numpy as np
import shap
import optuna
from catboost import *
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import  SVR
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score
shap.initjs()
from plotnine import *
```

Functions

[ ]:
```
### Importing datasets ###
def Training(file):
    BASE_DIR = r'training CSV file path'
```

```
    if os.path.exists(BASE_DIR):
        DATA_DIR = os.path.join(BASE_DIR,
                                "Training")  # use this if the data is inside␣
↪another folder inside the BASE_DIR
        data_file_name = file
        file_path = os.path.join(DATA_DIR, data_file_name)
        df = pd.read_csv(file_path)
    else:
        BASE_DIR = r'Alternative path'
        DATA_DIR = os.path.join(BASE_DIR,
                                "Training")  # use this if the data is inside␣
↪another folder inside the BASE_DIR
        data_file_name = file
        file_path = os.path.join(DATA_DIR, data_file_name)
        df = pd.read_csv(file_path)

    return df

def Testing(file):
    BASE_DIR = r'Testing CSV file path'
    if os.path.exists(BASE_DIR):
        DATA_DIR = os.path.join(BASE_DIR,
                                "Testing")  # use this if the data is inside␣
↪another folder inside the BASE_DIR
        data_file_name = file
        file_path = os.path.join(DATA_DIR, data_file_name)
        df = pd.read_csv(file_path)
    else:
        BASE_DIR = r'Alternative path'
        DATA_DIR = os.path.join(BASE_DIR,
                                "Testing")  # use this if the data is inside␣
↪another folder inside the BASE_DIR
        data_file_name = file
        file_path = os.path.join(DATA_DIR, data_file_name)
        df = pd.read_csv(file_path)

    return df
```

```
### Creating features ###
def Creating_Volume_Mass_Compaction(list):
    # Separating into Stone, sand, filler and bitumen % v/v
    list['stone_volume_percentage'] = list['Volume_Target_C22_4'] +␣
↪list['Volume_Target_C16'] + list[
        'Volume_Target_C11_2'] + list['Volume_Target_C08'] +␣
↪list['Volume_Target_C05_6'] + list['Volume_Target_C002mm']
    list['sand_volume_percentage'] = list['Volume_Target_C00063mu']
    list['filler_volume_percentage'] = list['Volume_Target_filler']

    # Separating into Stone, sand, filler and bitumen % m/m
    list['stone_mass_percentage'] = list['Mass_Target_C22_4'] +␣
↪list['Mass_Target_C16'] + list[
        'Mass_Target_C11_2'] + list['Mass_Target_C08'] + list['Mass_Target_C05_6']␣
↪+ list['Mass_Target_C002mm']
    list['sand_mass_percentage'] = list['Mass_Target_C00063mu']
    list['filler_mass_percentage'] = list['Mass_Target_filler']

    # Changing names of columns
    list.columns = list.columns.str.replace("_Target", "")
    list.columns = list.columns.str.replace("bit2_delta", "Phase Angle")
    list['year'] = list['year'].str.replace("yr", "")
    # list['Phase_Year'] = list['phase'] + '_' + list['year'] #Use this for Phase␣
↪analysis
    # Creating Compaction Degree
    list['compaction degree'] = list['densities'] / list['target_density'] * 100
    return list


### Separating independent features ###
def Group_Mass_stiffness(list):
    X = list.drop(
        ['stiffness', 'VA', 'filename', 'phase_tot', 'work', 'monsternames',␣
↪'densities', 'bit2_TRenK',
         'bit2_Gstar', 'percentage_bit', 'Volume_C22_4', 'Volume_C16',␣
↪'Volume_C11_2', 'Volume_C08',
         'Volume_C05_6', 'Volume_C002mm',
         'Volume_C00063mu', 'Volume_filler', 'Mass_C22_4', 'Mass_C16', 'Mass_C11_2',
```

```
        'Mass_C08', 'Mass_C05_6', 'Mass_C002mm', 'Mass_C00063mu', 'Mass_filler',␣
↪'stone_volume_percentage',
        'sand_volume_percentage', 'filler_volume_percentage', 'lab', 'phase',␣
↪'Volume_bitumen'], axis=1)
    return X


### Defining the type of feature ###
def Stiffness(list):
    categorical_data = ['phase', 'lab', 'mix_setup', 'comp_setup']

    numerical_data = ['year', 'densities', 'VA', 'bit2_pen', 'Phase Angle',␣
↪'target_density', 'percentage_PR',
                      'percentage_bit', 'Volume_C22_4', 'Volume_C16',␣
↪'Volume_C11_2', 'Volume_C08', 'Volume_C05_6',
                      'Volume_C002mm',
                      'Volume_C00063mu', 'Volume_filler', 'Volume_bitumen',␣
↪'Mass_C22_4', 'Mass_C16', 'Mass_C11_2',
                      'Mass_C08', 'Mass_C05_6', 'Mass_C002mm', 'Mass_C00063mu',␣
↪'Mass_filler', 'Mass_bitumen',
                      'stone_volume_percentage', 'sand_volume_percentage',␣
↪'filler_volume_percentage',
                      'stone_mass_percentage', 'sand_mass_percentage',␣
↪'filler_mass_percentage',
                      'stiffness']
    list[categorical_data] = list[categorical_data].astype(str)
    list[numerical_data] = list[numerical_data].astype('float')


### Plotting functions ###
def chart_regression(pred, y,Title, sort=True):
    t = pd.DataFrame({'pred': pred, 'y': y})
    if sort:
        t.sort_values(by=['y'], inplace=True)
    plt.plot(t['y'].tolist(), label='expected')
    plt.plot(t['pred'].tolist(), label='prediction')
    plt.ylabel(Title +'Regression Chart')
    plt.xlabel('0-100% of dataset')
```

```
    plt.legend()
    plt.show()

def chart_regression_SVM(pred, y, sort=True):
    t = pd.DataFrame({'pred': pred, 'y': y.flatten()})
    if sort:
        t.sort_values(by=['y'], inplace=True)
    plt.plot(t['y'].tolist(), label='expected')
    plt.plot(t['pred'].tolist(), label='prediction')
    plt.ylabel('Regression Chart')
    plt.xlabel('0-100% of dataset')
    plt.legend()
    plt.show()

def mix_scatter(y_train, y_pred_train, y_test, y_pred_test, r2_train, r2_test,␣
↪Title):
    data_val = pd.DataFrame({'y_truth': y_train, 'y_pred': y_pred_train, 'subset':␣
↪'Training and Validation Data'})
    data_test = pd.DataFrame({'y_truth': y_test, 'y_pred': y_pred_test, 'subset':␣
↪'Test Data'})
    data = data_val.append(data_test, sort=False)
    text_val = pd.DataFrame({'r2': [r'\rm R^2= {:.2f} % '.format(r2_train)]})
    text_test = pd.DataFrame({'r3': [r'\rm R^2= {:.2f} % '.format(r2_test)]})
    p = ggplot(data=data, mapping=aes(x='y_truth', y='y_pred'))
    p = p + geom_point(data=data_val, mapping=aes(x='y_truth', y='y_pred',␣
↪color='subset'), alpha=0.7, stroke=0, size=3)
    p = p + geom_point(data=data_test, mapping=aes(x='y_truth', y='y_pred',␣
↪color='subset'), alpha=0.7, stroke=0,
                       size=3)
    p = p + geom_line(mapping=aes(y='y_truth'), color='black')
    p = p + theme(axis_text=element_text(size=10), axis_title=element_text(size=10))
    p = p + labs(title=Title)
    p = p + xlab('Measured ')
    p = p + ylab('Predicted ')
    p = p + scale_color_manual(values=["#F8766D", "#00BFC4", 'blue'])
    p = p + geom_text(data=text_val, mapping=aes(x=6800, y=18000, label='r2'),␣
↪color="#F8766D", nudge_x=-0.1, parse=True,
                      family='Palatino')
```

```
    p = p + geom_text(data=text_test, mapping=aes(x=6800, y=17250, label='r3'),␣
↪color="#00BFC4", nudge_x=-0.1, parse=True,
                      family='Palatino')
    print(p)

### Prediction + plotting function ###
def pred(model, X_train, y_train, X_test, y_test, Title):

    y_pred_train = model.predict(X_train)
    y_pred_test = model.predict(X_test)

    r2_train = r2_score(y_train, y_pred_train)
    r2_test = r2_score(y_test, y_pred_test)

    print("Rš with validation data ", r2_train)
    print("Rš with Test data ", r2_test)

    mix_scatter(y_train, y_pred_train, y_test, y_pred_test, r2_train, r2_test,␣
↪Title)
    chart_regression(y_pred_train, y_train, Title)

def pred_SVR(model, scaler_y, X_train, y_train, X_test, y_test, Title):

    y_pred_train = scaler_y.inverse_transform([model.predict(X_train)])
    y_pred_test = scaler_y.inverse_transform([model.predict(X_test)])

    y_pred_train = y_pred_train.reshape(-1, 1)
    y_pred_test = y_pred_test.reshape(-1, 1)

    y_train =  scaler_y.inverse_transform(y_train)
    y_test =  scaler_y.inverse_transform(y_test)

    r2_training = r2_score(y_train, y_pred_train)
    r2_test = r2_score(y_test, y_pred_test)

    print("Rš with training data ", r2_training)
    print("Rš with test data ", r2_test)
```

```
    chart_regression_SVM(y_pred_train.flatten(), y_train, Title)
    mix_scatter(y_train.flatten(), y_pred_train.flatten(), y_test.flatten(),␣
↪y_pred_test.flatten(),  r2_training, r2_test, Title)
```

Importing the dataset

[ ]:
```
file = 'Train_stiffness.csv'
file_test = 'Test_stiffness.csv'

df = Training(file)
df_test = Testing(file_test)
```

Data Preprocessing

[ ]:
```
df = Creating_Volume_Mass_Compaction(df)
df_test = Creating_Volume_Mass_Compaction(df_test)
```

[ ]:
```
categorical_data = ['mix_setup', 'comp_setup']
Stiffness(df)
Stiffness(df_test)
```

[ ]:
```
# This step one-hot-encode the categorical data. Only for SVR and RF
df_RF_SVR = pd.get_dummies(df, columns=categorical_data)
df_test_RF_SVR = pd.get_dummies(df_test, columns=categorical_data)
```

[ ]:
```
# Independent Training variables
X = Group_Mass_stiffness(df)
X_test = Group_Mass_stiffness(df_test)

# Dependent Training variables
y = df.stiffness
y_test = df_test.stiffness
```

[ ]:
```
# Separating into dependent and independent variables for encoded data. Only for RF␣
↪and SVR

# Independent Training variables
X_RF = Group_Mass_stiffness(df_RF_SVR)
X_test_RF = Group_Mass_stiffness(df_test_RF_SVR)

# Dependent Training variables
y_RF = df_RF_SVR.stiffness
y_test_RF = df_test_RF_SVR.stiffness
```

[ ]:
```
# After the split the data should be scaled. Only required for SVR. The encoded data should not be scaled.

sc_x = StandardScaler()
sc_y = StandardScaler()

X_SVR = X_RF
X_test_SVR = X_test_RF

X_SVR.iloc[:, 0:9] = sc_x.fit_transform(X_SVR.iloc[:, 0:9])
X_test_SVR.iloc[:, 0:9] = sc_x.transform(X_test_SVR.iloc[:, 0:9])

y_SVR = y_RF
y_test_SVR = y_test_RF

# This step transform the data into 1D array. Only required for SVR
y_SVR = y_SVR.values.reshape(-1, 1)
y_test_SVR = y_test_RF.values.reshape(-1, 1)

y_SVR = sc_y.fit_transform(y_SVR)
y_test_SVR = sc_y.transform(y_test_SVR)
```

Splitting Data

[ ]:
```
'''
This data was originally split in the beforehand. This step is required if the data still have to be split or further splitting/cross-validation required.

'''
# X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=1237)
```

[ ]:
```
### Test and training pool for catboost ###
Pool_train = Pool(X, y, cat_features=categorical_data)
Pool_test = Pool(X_test, y_test, cat_features=categorical_data)
```

Fitting

[ ]:
```
Model_GB = CatBoostRegressor()
Model_RF = RandomForestRegressor()
Model_SVR = SVR()
```

```
Model_GB.fit(Pool_train, eval_set=Pool_test)
Model_RF.fit(X_RF,y_RF)
Model_SVR.fit(X_SVR,y_SVR)
```

Predicting

[ ]:
```
pred(Model_GB,X,y,X_test,y_test, 'Gradient Boosting')
```

[ ]:
```
pred(Model_RF,X_RF,y_RF,X_test_RF,y_test_RF, 'Random Forests')
```

[ ]:
```
pred_SVR(Model_SVR,sc_y,X_SVR,y_SVR,X_test_SVR,y_test_SVR, 'Support Vector Machine')
```

E.2 Optimization and cross-validation

Importing libraries

[ ]:
```
from catboost import *
import optuna
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.metrics import r2_score, mean_squared_error
```

Optimization function

[ ]:
```
def Optuna_grid(X, y, categorical_data, title):


    X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2,␣
↪random_state=1861)


    train_pool = Pool(X_train, y_train, cat_features=categorical_data)
    val_pool = Pool(X_val, y_val, cat_features=categorical_data)


    def objective(trial):
        param = {}
        param['learning_rate'] = trial.suggest_discrete_uniform("learning_rate", 0.
↪001, 0.02, 0.001)
        param['depth'] = trial.suggest_int('depth', 9, 15)
        param['iterations'] = 10000
        param['use_best_model'] = True
        param['eval_metric'] = 'RMSE'
        param['od_type'] = 'Iter'
        param['od_wait'] = 20
        param['logging_level'] = 'Silent'
```

```
        model = CatBoostRegressor(**param, loss_function='RMSE')

        model.fit(train_pool,
                  eval_set=val_pool)

        y_pred = model.predict(X_val)
        score = model.score(y_pred,y_val)
        return score

    study = optuna.create_study(study_name=f'Catboost_{title}',␣
↪direction="minimize")
    study.optimize(objective, n_trials=100)

    trial = study.best_trial

    print("Number of completed trials: {}".format(len(study.trials)))
    print("Best trial:")

    print("\tBest Score: {}".format(trial.value))
    print("\tBest Params: ")
    for key, value in trial.params.items():
        print("    {}: {}".format(key, value))

    optuna.visualization.matplotlib.plot_param_importances(study)

    return study
```

Cross-validation function

```
[ ]: def cross_val(X,y,categorical_data,study,):

    kf = KFold(n_splits=10, shuffle=True, random_state=216)
    models = []
    scores = []
    error_train = []
    error_val = []

    for train_index, val_index in kf.split(X, y):

        X_train, X_val = X.iloc[train_index], X.iloc[val_index]
```

```
        y_train, y_val = y.iloc[train_index], y.iloc[val_index]

        train_pool = Pool(X_train, y_train, cat_features=categorical_data)
        val_pool = Pool(X_val, y_val, cat_features=categorical_data)

        optimized_regressor = CatBoostRegressor(learning_rate=study.
↪best_params['learning_rate'],
                                                depth=study.best_params['depth'],
                                                iterations=10000,
                                                use_best_model=True,
                                                loss_function='RMSE',
                                                od_type='iter',
                                                od_wait=20,
                                                random_state=13,
                                                logging_level='Silent')
        optimized_regressor.fit(train_pool,
                                eval_set=val_pool)

        pred_val = optimized_regressor.predict(X_val)

        error_train.append(mean_squared_error(y_train, optimized_regressor.
↪predict(X_train), squared=False))
        error_val.append(mean_squared_error(y_val, optimized_regressor.
↪predict(X_val), squared=False))

        scores.append(r2_score(y_val,pred_val))
        models.append(optimized_regressor)

    models_avrg = sum_models(models,
                             weights=[1.0/len(models)] * len(models))

    return models_avrg
```

E.3 API

The machine learning analysis was programmed in *Python* with the assistance of several libraries. A small summary of the used application programming interface (API) is represented in Figure E.1.

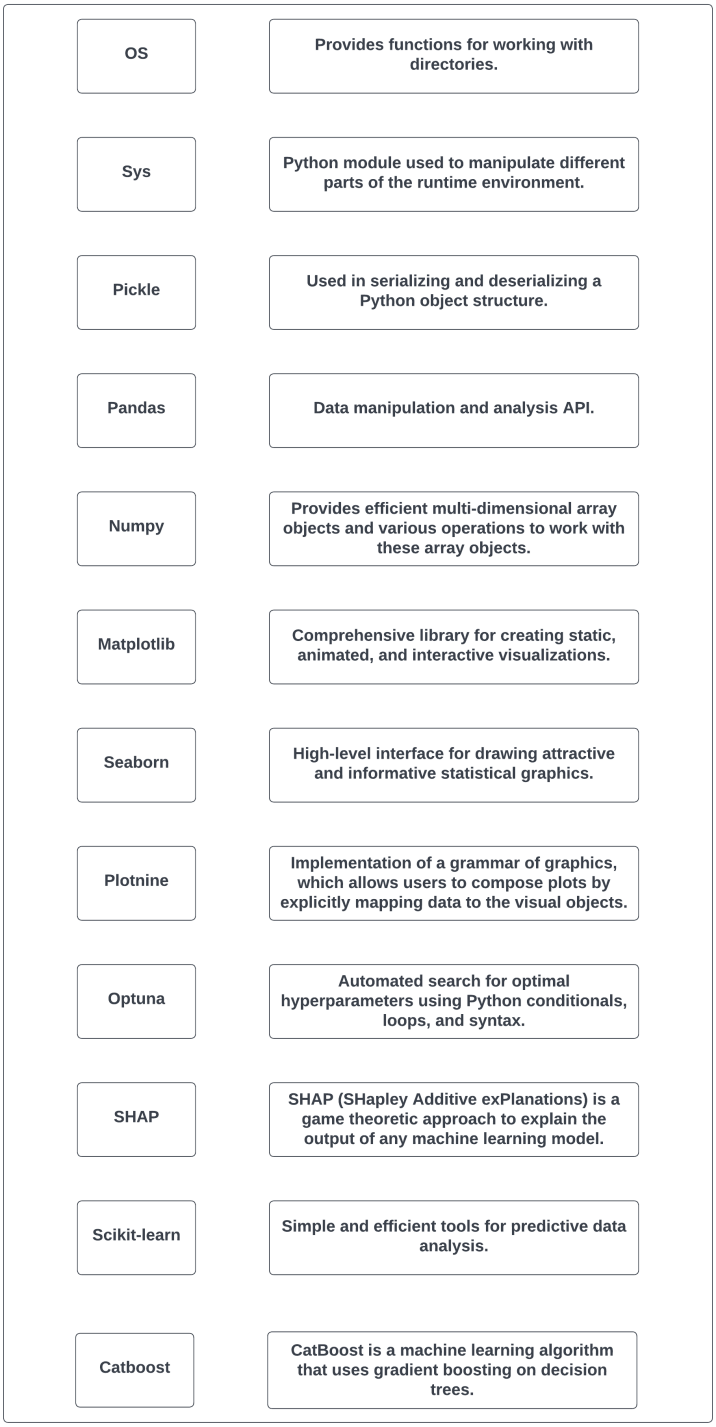


Figure E.1: Libraries used in the research