Inferring the residential building type from 3DBAG

Master Thesis - P5

Hoi-Kang Poon (Chris)
Student #4355938

1st supervisor: Camilo León-Sánchez
2nd supervisor: Giorgio Agugiaro
Co-reader: Nail Ibrahimli

Introduction

- Urban energy consumption

- Urban Energy Modeling

  Better energy management
  Cities' design optimization
  Inaccurate understanding of energy use
  Missed opportunities for energy savings

Requires:

- 3DBAG

- IEE project TABULA

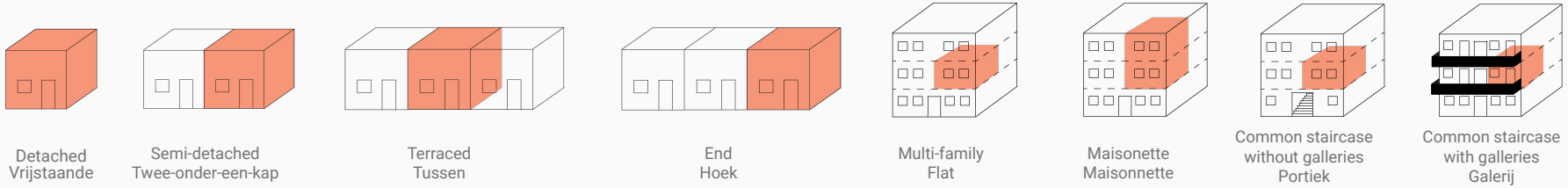
Research questions

- To what extent can machine learning correctly classify the building stock of the Netherlands?
  - What **features** are needed to infer the building types of the buildings of the 3DBAG?
  - Which **data** are required?
  - Which (combination of) **machine learning** algorithms is the *most suitable* to be used for the classification of the building stock of the Netherlands, with regards to the size and nature of the data used, the availability of computational resources, the interpretability of the results and the desired level of accuracy?

Contents

Theoretical background and related work

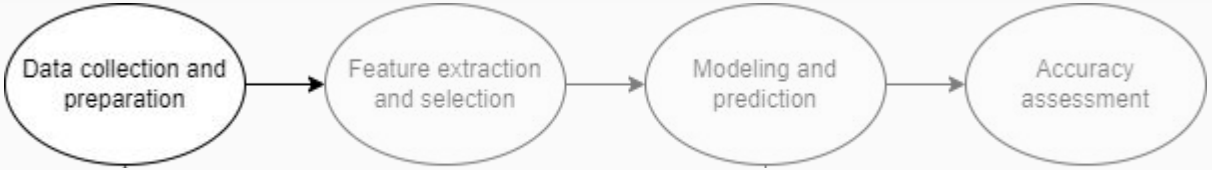
- Dutch residential buildings



- Machine learning
  - **Supervised learning** - Unsupervised learning
  - **Labelled data** - Unlabelled data
    - = features + desired outcome
  - **Classification** + Regression - Clustering
    - = predict class labels → building types

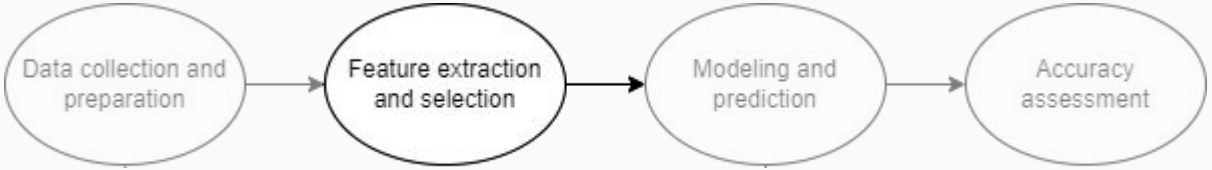
- 3D building metrics (or features)
  - Measurable property or characteristics (of a building)
  - For example:
    - Footprint area and perimeter
    - Height, width, length and volume
    - Wall and roof areas

Method



- 3DBAG → **building stock geometry**
  - 3D building models of the building stock of the Netherlands
- BAG → **cadastral features**
  - National cadastral dataset
- Rijssen-Holten → **ground truth (labelled data)**
  - Open testbed for energy applications, study area is located in municipality of Rijssen-Holten
- EP-online → **ground truth (preprocessing needed)**
  - Official national database containing energy labels and energy performance indicators of buildings

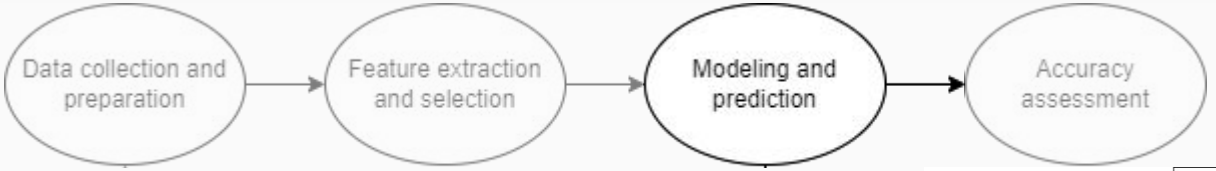
Method



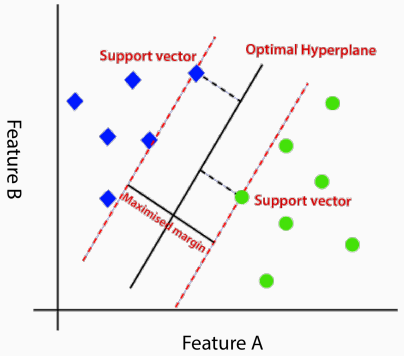
	Features	Source	Details and relevance
1	Number of adjacent buildings	BAG	See classification process of the Kadaster
2	Number of adjacent buildings of adjacent buildings.	BAG	See classification process of the Kadaster
3	Year of construction	BAG	Construction year might not be relevant to inferring the residential building type, but it is relevant to determining the construction year class of a residential building type.
4	Number of dwellings in the building	BAG	See classification process of the Kadaster
5	Footprint area	BAG	Describes the building geometry.
6	Footprint perimeter	BAG	Describes the building geometry and provides additional information about the footprint shape, like the compactness and complexity.
7	Number of the vertices in footprint	BAG	The number of the vertices gives another indication of the complexity of the footprint shape.
8	The number of neighbouring building (radius: 25m)	BAG	The number of neighbouring building centroids within a certain radius of the footprint centroid. For example, taller buildings, like apartment blocks generally have more open space in the surroundings.
9	The number of neighbouring building (radius: 50m)	BAG	The same as above, but larger radius
10	The number of neighbouring building (radius: 75m)	BAG	The same as above, but larger radius
11	The number of neighbouring building (radius: 100m)	BAG	The same as above, but larger radius
12	Actual volume in LoD1.2	3DBAG	Describes building geometry, utilising LoD1.2 of the 3D BAG
13	Convex hull volume in LoD1.2	3DBAG	Describes building geometry, utilising LoD1.2 of the 3D BAG
14	Oriented bounding box width in LoD1.2	3DBAG	Describes building geometry, utilising LoD1.2 of the 3D BAG
15	Oriented bounding box length in LoD1.2	3DBAG	Describes building geometry, utilising LoD1.2 of the 3D BAG
16	Total wall surface area in LoD1.2	3DBAG	Describes building geometry, utilising LoD1.2 of the 3D BAG
17	Total roof surface area in LoD1.2	3DBAG	Describes building geometry, utilising LoD1.2 of the 3D BAG
18	Maximum height in LoD1.2	3DBAG	Describes building geometry, utilising LoD1.2 of the 3D BAG
19	Actual volume in LoD2.2	3DBAG	Describes building geometry, utilising LoD1.2 of the 3D BAG
20	Convex hull volume in LoD2.2	3DBAG	Describes building geometry, utilising LoD2.2 of the 3D BAG
21	Total wall surface area in LoD2.2	3DBAG	Describes building geometry, utilising LoD2.2 of the 3D BAG
22	Total roof surface area in LoD2.2	3DBAG	Describes building geometry, utilising LoD2.2 of the 3D BAG
23	Maximum height in LoD2.2	3DBAG	Describes building geometry, utilising LoD2.2 of the 3D BAG
24	Height (without roof) in LoD2.2	3DBAG	Describes building geometry, utilising LoD2.2 of the 3D BAG
25	*Number of storeys of the building*	*Rijssen-Holten*	*Number of storeys of the building*

- 25 features extracted
- Selecting 7, based on:
  - ANOVA-F
  - Mutual Information
  - Impurity
  - Permutation

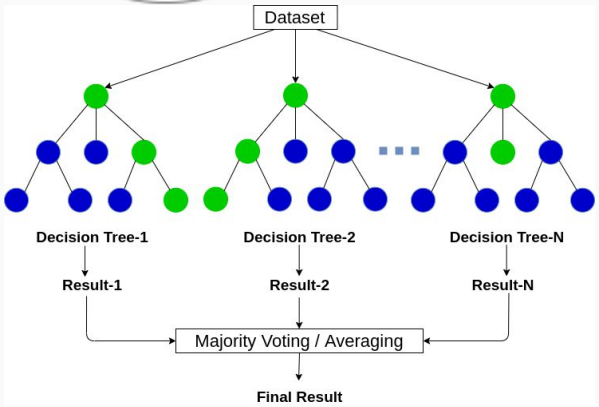
Method



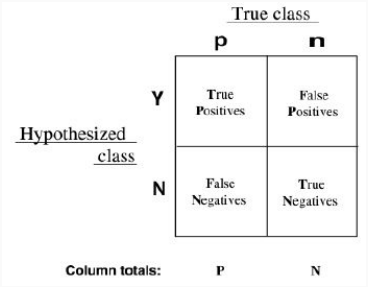
- Support Vector Machine



- Random Forest



- Confusion Matrix
  - Precision
  - Recall
  - F1-score

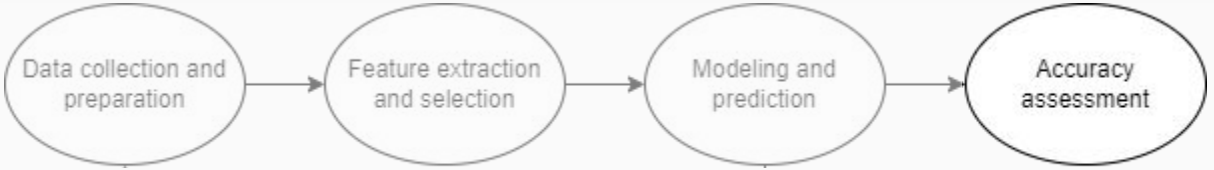


$$\text{Precision} = \frac{TP}{(TP + FP)}$$

$$\text{Recall} = \frac{TP}{(TP + FN)}$$

$$\text{F1-score} = \frac{2(Precision \cdot Recall)}{(Precision + Recall)}$$

Method



- Model application
- Hit-and-miss analysis
  - For example:

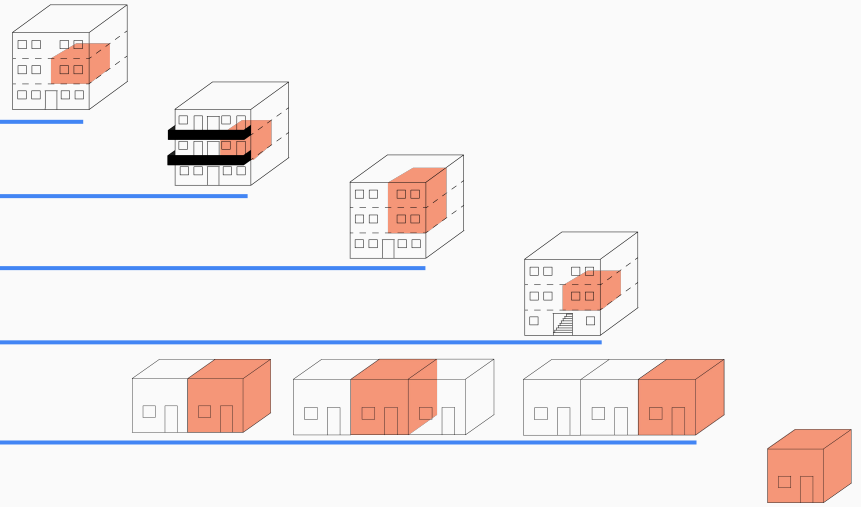
Label	Prediction	In Real Life	
TH	TH	TH	HIT!
TH	TH	SDH	MISS!
TH	SDH	SDH	HIT!

Implementation: Case studies

#	Location	Usage
1	Rijssen-Holten	Train model
2	Delft	Train model
3	Duivendrecht	Focus on: **MFH**
4	Bijlmer-Oost	Focus on: **Galerijwoning**
5	Borneo-Sporenburg	Focus on: **Maisonette**
6	Laakkwartier	Focus on: **Portiekwoning**
7	Oud-Diemen	Focus on: **SDH, TH, EH**
8	Laren	Focus on: **DH**

Building type	Name	Count	Percentage
Galerijwoning	Common staircase with galleries	4	0.06%
Maisonnettewoning	Maisonettes	4	0.06%
Portiekwoning	Common staircase without galleries	8	0.11%
Flatwoning	Multi-family house	58	0.80%
Hoekwoning	End house	1452	20.06%
Vrijstaande Woning	Detached	1783	24.63%
Twee-onder-een-kapwoning	Semi-detached	1785	24.66%
Tussenwoning	Terraced house	2145	29.63%
TOTAL		**7239**	**100.00%**

Building type	Name	Count	Percentage
Galerijwoning	Common staircase with galleries	63	0.60%
Portiekwoning	Common staircase without galleries	118	1.12%
Vrijstaande Woning	Detached	253	2.39%
Twee-onder-een-kapwoning	Semi-detached	552	5.22%
Maisonnettewoning	Maisonettes	681	6.44%
Hoekwoning	End house	1437	13.58%
Flatwoning	Multi-family house	1476	13.95%
Tussenwoning	Terraced house	5999	56.71%
TOTAL		**10579**	**100.00%**

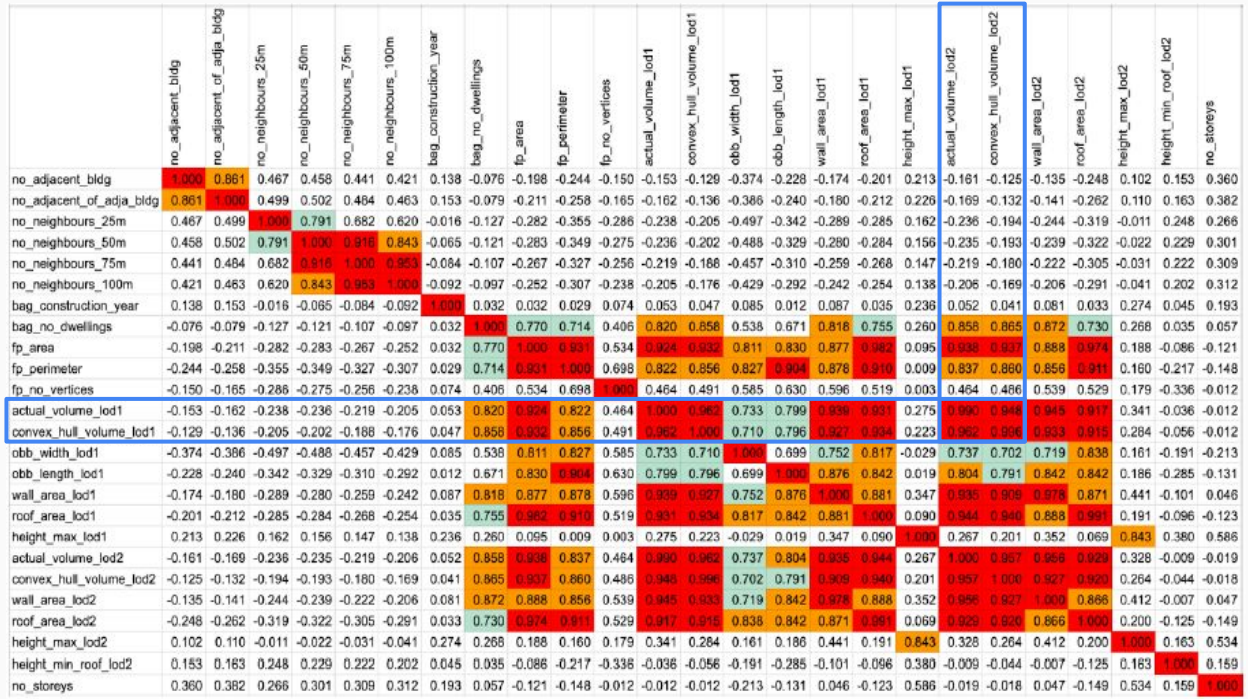


Implementation: Feature engineering

- Analysis

Removed based on correlation:

- Actual volume in LoD1
- C. hull volume in LoD1
- C. hull volume in LoD2
- Wall area in LoD1
- Roof area in LoD1
- Maximum height in LoD1



	no_adjacent_bldg	no_adjacent_of_adja_bldg	no_neighbours_25m	no_neighbours_50m	no_neighbours_75m	no_neighbours_100m	bag_construction_year	bag_no_dwellings	fp_area	fp_perimeter	fp_no_vertices	actual_volume_lod1	convex_hull_volume_lod1	obb_width_lod1	obb_length_lod1	wall_area_lod1	roof_area_lod1	height_max_lod1	actual_volume_lod2	convex_hull_volume_lod2	wall_area_lod2	roof_area_lod2	height_max_lod2	height_min_roof_lod2	no_storeys
no_adjacent_bldg	1.000	0.861	0.467	0.458	0.441	0.421	0.138	-0.076	-0.198	-0.244	-0.150	-0.153	-0.129	-0.374	-0.228	-0.174	-0.201	0.213	-0.161	-0.125	-0.135	-0.248	0.102	0.153	0.360
no_adjacent_of_adja_bldg	0.861	1.000	0.499	0.502	0.484	0.463	0.153	-0.079	-0.211	-0.258	-0.165	-0.162	-0.136	-0.386	-0.240	-0.180	-0.212	0.226	-0.169	-0.132	-0.141	-0.262	0.110	0.163	0.382
no_neighbours_25m	0.467	0.499	1.000	0.791	0.682	0.620	-0.016	-0.127	-0.282	-0.355	-0.286	-0.238	-0.205	-0.497	-0.342	-0.289	-0.285	0.162	-0.236	-0.194	-0.244	-0.319	-0.011	0.248	0.266
no_neighbours_50m	0.458	0.502	0.791	1.000	0.916	0.843	-0.065	-0.121	-0.283	-0.349	-0.275	-0.236	-0.202	-0.488	-0.329	-0.280	-0.284	0.156	-0.235	-0.193	-0.239	-0.322	-0.022	0.229	0.301
no_neighbours_75m	0.441	0.484	0.682	0.916	1.000	0.953	-0.084	-0.107	-0.267	-0.327	-0.256	-0.219	-0.188	-0.457	-0.310	-0.259	-0.268	0.147	-0.219	-0.180	-0.222	-0.305	-0.031	0.222	0.309
no_neighbours_100m	0.421	0.463	0.620	0.843	0.953	1.000	-0.092	-0.097	-0.252	-0.307	-0.238	-0.205	-0.176	-0.429	-0.292	-0.242	-0.254	0.138	-0.206	-0.169	-0.206	-0.291	-0.041	0.202	0.312
bag_construction_year	0.138	0.153	-0.016	-0.065	-0.084	-0.092	1.000	0.032	0.032	0.029	0.074	0.053	0.047	0.085	0.012	0.087	0.035	0.236	0.052	0.041	0.081	0.033	0.274	0.045	0.193
bag_no_dwellings	-0.076	-0.079	-0.127	-0.121	-0.107	-0.097	0.032	1.000	0.770	0.714	0.406	0.820	0.858	0.538	0.671	0.816	0.755	0.260	0.858	0.865	0.872	0.730	0.268	0.035	0.057
fp_area	-0.198	-0.211	-0.282	-0.283	-0.267	-0.252	0.032	0.770	1.000	0.931	0.534	0.924	0.932	0.811	0.830	0.877	0.982	0.095	0.938	0.937	0.888	0.974	0.188	-0.086	-0.121
fp_perimeter	-0.244	-0.258	-0.355	-0.349	-0.327	-0.307	0.029	0.714	0.931	1.000	0.698	0.822	0.856	0.827	0.904	0.878	0.910	0.009	0.837	0.860	0.856	0.911	0.160	-0.217	-0.148
fp_no_vertices	-0.150	-0.165	-0.286	-0.275	-0.256	-0.238	0.074	0.406	0.534	0.698	1.000	0.464	0.491	0.585	0.630	0.596	0.519	0.003	0.464	0.486	0.539	0.529	0.179	-0.336	-0.012
actual_volume_lod1	-0.153	-0.162	-0.238	-0.236	-0.219	-0.205	0.053	0.820	0.924	0.822	0.464	1.000	0.962	0.733	0.799	0.939	0.931	0.275	0.990	0.948	0.945	0.917	0.341	-0.036	-0.012
convex_hull_volume_lod1	-0.129	-0.136	-0.205	-0.202	-0.188	-0.176	0.047	0.858	0.932	0.856	0.491	0.962	1.000	0.710	0.796	0.927	0.934	0.223	0.962	0.996	0.933	0.915	0.284	-0.056	-0.012
obb_width_lod1	-0.374	-0.386	-0.497	-0.488	-0.457	-0.429	0.085	0.538	0.811	0.827	0.585	0.733	0.710	1.000	0.699	0.752	0.817	-0.029	0.737	0.702	0.719	0.838	0.161	-0.191	-0.213
obb_length_lod1	-0.228	-0.240	-0.342	-0.329	-0.310	-0.292	0.012	0.671	0.830	0.904	0.630	0.799	0.796	0.699	1.000	0.876	0.842	0.019	0.804	0.791	0.842	0.842	0.186	-0.285	-0.131
wall_area_lod1	-0.174	-0.180	-0.289	-0.280	-0.259	-0.242	0.087	0.818	0.877	0.878	0.596	0.939	0.927	0.752	0.876	1.000	0.881	0.347	0.935	0.909	0.978	0.871	0.441	-0.101	0.046
roof_area_lod1	-0.201	-0.212	-0.285	-0.284	-0.268	-0.254	0.035	0.755	0.982	0.910	0.519	0.931	0.934	0.817	0.842	0.881	1.000	0.090	0.944	0.940	0.888	0.991	0.191	-0.096	-0.123
height_max_lod1	0.213	0.226	0.162	0.156	0.147	0.138	0.236	0.260	0.095	0.009	0.003	0.275	0.223	-0.029	0.019	0.347	0.090	1.000	0.267	0.201	0.352	0.069	0.843	0.380	0.586
actual_volume_lod2	-0.161	-0.169	-0.236	-0.235	-0.219	-0.206	0.052	0.858	0.938	0.837	0.464	0.990	0.962	0.737	0.804	0.935	0.944	0.267	1.000	0.957	0.956	0.929	0.328	-0.009	-0.019
convex_hull_volume_lod2	-0.125	-0.132	-0.194	-0.193	-0.180	-0.169	0.041	0.865	0.937	0.860	0.486	0.948	0.996	0.702	0.791	0.909	0.949	0.201	0.957	1.000	0.927	0.920	0.264	-0.044	-0.018
wall_area_lod2	-0.135	-0.141	-0.244	-0.239	-0.222	-0.206	0.081	0.872	0.888	0.856	0.539	0.945	0.933	0.719	0.842	0.978	0.888	0.352	0.956	0.927	1.000	0.866	0.412	-0.007	0.047
roof_area_lod2	-0.248	-0.262	-0.319	-0.322	-0.305	-0.291	0.033	0.730	0.974	0.911	0.529	0.917	0.915	0.838	0.842	0.871	0.991	0.069	0.929	0.920	0.866	1.000	0.200	-0.125	-0.149
height_max_lod2	0.102	0.110	-0.011	-0.022	-0.031	-0.041	0.274	0.268	0.188	0.160	0.179	0.341	0.264	0.161	0.186	0.441	0.191	0.843	0.328	0.264	0.412	0.200	1.000	0.163	0.534
height_min_roof_lod2	0.153	0.163	0.248	0.229	0.222	0.202	0.045	0.035	-0.086	-0.217	-0.336	-0.036	-0.056	-0.191	-0.285	-0.101	-0.096	0.380	-0.009	-0.044	-0.007	-0.125	0.163	1.000	0.159
no_storeys	0.360	0.382	0.266	0.301	0.309	0.312	0.193	0.057	-0.121	-0.148	-0.012	-0.012	-0.012	-0.213	-0.131	0.046	-0.123	0.586	-0.019	-0.018	0.047	-0.149	0.534	0.159	1.000

- Selection

Implementation: Feature engineering

- Analysis
- Selection

Most used:

- Number of adjacent buildings
- Number of adjacent buildings of adjacent buildings
- Number of dwellings in the building
- Oriented bounding box width in LoD1

		Case Study 1				Case Study 2				
		SVM ANOVA-F	SVM MI	RF impurity	RF permu.	SVM ANOVA-F	SVM MI	RF impurity	RF permu.	Count
1	Number of adjacent buildings	x	x	x	x	x		x	x	7
2	Number of adjacent buildings of adjacent buildings.	x	x	x	x			x	x	6
3	Year of construction		x	x	x					3
4	Number of dwellings in the building	x			x	x	x	x	x	6
5	Footprint area		x				x	x	x	4
6	Footprint perimeter					x	x			2
7	Number of the vertices in footprint									0
8	The number of neighbouring building (radius: 25m)			x	x					2
9	The number of neighbouring building (radius: 50m)	x	x							2
10	The number of neighbouring building (radius: 75m)									0
11	The number of neighbouring building (radius: 100m)									0
12	Oriented bounding box width in LoD1.2	x	x	x	x	x	x	x	x	8
13	Oriented bounding box length in LoD1.2					x	x		x	3
14	Actual volume in LoD2.2	x	x	x			x	x		5
15	Total wall surface area in LoD2.2	x			x	x			x	4
16	Total roof surface area in LoD2.2						x			1
17	Maximum height in LoD2.2			x		x		x		3
18	Height (without roof) in LoD2.2									0

Results

- Case study 1 models:
  (Test split 20%)

Model	Tuning time (s)	Training time (s)	Accuracy	Balanced accuracy
SVM ANOVA-F	25.2	1.42	0.918	0.583
SVM MI	20.43	0.35	0.917	0.571
RF impurity	426.31	2.87	0.911	0.741
RF permutation	344.27	2.48	0.918	0.652

- Case study 2 models:
  (Test split 20%)

Model	Tuning time (s)	Training time (s)	Accuracy	Balanced accuracy
SVM ANOVA-F	112.9	7.04	0.672	0.395
SVM MI	240.39	8.75	0.389	0.330
RF impurity	996.91	14.16	0.711	0.737
RF permutation	1105.94	17.31	0.733	0.737

- SVM is faster than RF
- C1 models are faster, because smaller (input) dataset
- Overall high accuracy, but lower balanced accuracy

Results

- Model comparison:
  - Higher accuracies in C1
  - Even C2 on test split
  - While C2 has better class balance

		Case Study 1				Case Study 2			
		SVM ANOVA-F	**SVM MI**	**RF impurity**	**RF permu.**	**SVM ANOVA-F**	**SVM MI**	**RF impurity**	**RF permu.**
Test split (20%)	Accuracy	91.8%	91.7%	91.1%	91.8%	67.2%	38.9%	71.1%	73.3%
	Balanced accuracy	58.3%	57.1%	74.1%	65.2%	39.5%	33.0%	73.7%	73.7%
On other case (whole dataset)	Accuracy	55.3%	55.6%	56.6%	56.1%	60.3%	38.3%	81.2%	84.0%
	Balanced accuracy	42.0%	40.4%	40.4%	44.7%	47.6%	23.8%	58.0%	62.5%

- Model application:
  - Accuracy and balanced accuracy higher

		Case Study 1				Case Study 2			
		SVM ANOVA-F	**SVM MI**	**RF impurity**	**RF permu.**	**SVM ANOVA-F**	**SVM MI**	**RF impurity**	**RF permu.**
Case Study 3 flat	Accuracy	75.8%	79.5%	87.9%	82.1%	85.4%	43.7%	80.2%	**90.4%**
	Balanced accuracy	49.2%	50.6%	53.9%	51.6%	36.7%	24.9%	**74.9%**	66.7%
Case Study 4 galerij	Accuracy	85.5%	86.0%	87.7%	87.3%	80.3%	30.0%	**90.3%**	89.4%
	Balanced accuracy	59.6%	60.4%	64.5%	60.5%	32.3%	11.3%	78.2%	**81.5%**
Case Study 5 maisonnette	Accuracy	60.1%	59.9%	63.5%	61.1%	**75.0%**	53.4%	67.5%	69.8%
	Balanced accuracy	54.7%	49.6%	62.3%	**65.5%**	39.6%	26.0%	60.5%	60.6%
Case Study 6 portiek	Accuracy	17.0%	18.1%	26.7%	63.1%	**82.1%**	34.0%	68.5%	71.2%
	Balanced accuracy	42.8%	43.0%	44.8%	51.6%	40.2%	33.0%	**68.6%**	62.6%
Case Study 7 rij_2o1k	Accuracy	**89.4%**	83.2%	89.2%	89.0%	60.8%	27.9%	80.2%	86.0%
	Balanced accuracy	71.8%	62.8%	67.7%	67.4%	35.4%	23.9%	74.9%	**78.9%**
Case Study 8 vrijstaande	Accuracy	98.0%	**98.5%**	81.9%	**98.5%**	86.9%	83.4%	93.0%	93.0%
	Balanced accuracy	90.6%	**94.2%**	66.6%	91.4%	43.4%	28.6%	72.0%	64.2%

Discussion

- Model comparison:
  - Classification reports
  - Effects of imbalanced class distribution

		Case Study 1				Case Study 2			
		SVM ANOVA-F	SVM MI	RF impurity	RF permu.	SVM ANOVA-F	SVM MI	RF impurity	RF permu.
Test split (20%)	Accuracy	91.8%	91.7%	91.1%	91.8%	67.2%	38.9%	71.1%	73.3%
	Balanced accuracy	58.3%	57.1%	74.1%	65.2%	39.5%	33.0%	73.7%	73.7%
On other case (whole dataset)	Accuracy	55.3%	55.6%	56.6%	56.1%	60.3%	38.3%	81.2%	84.0%
	Balanced accuracy	42.0%	40.4%	40.4%	44.7%	47.6%	23.8%	58.0%	62.5%

	precision	recall	f1-score	support
Flatwoning	0.59	0.91	0.71	11
Galerijwoning	0.00	0.00	0.00	1
Hoekwoning	0.93	0.87	0.90	287
Maisonnettewoning	0.00	0.00	0.00	1
Portiekwoning	0.00	0.00	0.00	1
Tussenwoning	0.94	0.94	0.94	427
Twee-onder-een-kap	0.88	0.89	0.88	353
Vrijstaande woning	0.94	0.96	0.95	353

	precision	recall	f1-score	support
Flatwoning	0.27	0.27	0.27	11
Galerijwoning	0.50	1.00	0.67	1
Hoekwoning	0.94	0.87	0.90	287
Maisonnettewoning	0.20	1.00	0.33	1
Portiekwoning	0.00	0.00	0.00	1
Tussenwoning	0.94	0.94	0.94	427
Twee-onder-een-kap	0.87	0.90	0.88	353
Vrijstaande woning	0.95	0.95	0.95	353

	precision	recall	f1-score	support
Flatwoning	0.26	0.46	0.34	295
Galerijwoning	0.02	0.08	0.03	13
Hoekwoning	0.00	0.00	0.00	287
Maisonnettewoning	0.01	0.01	0.01	136
Portiekwoning	0.10	0.71	0.17	24
Tussenwoning	0.78	0.52	0.62	1200
Twee-onder-een-kap	0.00	0.00	0.00	110
Vrijstaande woning	0.22	0.86	0.35	51

	precision	recall	f1-score	support
Flatwoning	0.27	0.27	0.27	295
Galerijwoning	0.50	1.00	0.67	13
Hoekwoning	0.94	0.87	0.90	287
Maisonnettewoning	0.20	1.00	0.33	136
Portiekwoning	0.00	0.00	0.00	24
Tussenwoning	0.94	0.94	0.94	1200
Twee-onder-een-kap	0.87	0.90	0.88	110
Vrijstaande woning	0.95	0.95	0.95	51

Discussion

- Model application:
  - C1 better on SDH, EH and TH
  - C2 better on MFH, galerij, maisonette and portiek
  - c2 SVM MI model performed worse overall
- F1-score:
  - How well a model can recognize a specific building type

	F1-score	Model
Case study 3 flat	0.95	c2 SVM ANOVA-F
Case study 4 galerij	0.69	**c2 RF impurity**, **c2 RF permutation**
Case study 5 maisonnette	0.30	**c2 RF permutation**
Case study 6 portiek	-	-
Case study 7 rij_2o1k	0.94*	c1 RF impurity, c1 RF permutation, c2 RF permutation
Case study 8 vrijstaande	1.00	**c1 SVM MI**, **c1 RF permutation**

		Case Study 1				Case Study 2			
		SVM ANOVA-F	**SVM MI**	**RF impurity**	**RF permu.**	**SVM ANOVA-F**	**SVM MI**	**RF impurity**	**RF permu.**
Case Study 3 flat	Accuracy	75.8%	79.5%	87.9%	82.1%	85.4%	43.7%	80.2%	**90.4%**
	Balanced accuracy	49.2%	50.6%	53.9%	51.6%	36.7%	24.9%	**74.9%**	66.7%
Case Study 4 galerij	Accuracy	85.5%	86.0%	87.7%	87.3%	80.3%	30.0%	**90.3%**	89.4%
	Balanced accuracy	59.6%	60.4%	64.5%	60.5%	32.3%	11.3%	78.2%	**81.5%**
Case Study 5 maisonnette	Accuracy	60.1%	59.9%	63.5%	61.1%	**75.0%**	53.4%	67.5%	69.8%
	Balanced accuracy	54.7%	49.6%	62.3%	**65.5%**	39.6%	26.0%	60.5%	60.6%
Case Study 6 portiek	Accuracy	17.0%	18.1%	26.7%	63.1%	**82.1%**	34.0%	68.5%	71.2%
	Balanced accuracy	42.8%	43.0%	44.8%	51.6%	40.2%	33.0%	**68.6%**	62.6%
Case Study 7 rij_2o1k	Accuracy	**89.4%**	83.2%	89.2%	89.0%	60.8%	27.9%	80.2%	86.0%
	Balanced accuracy	71.8%	62.8%	67.7%	67.4%	35.4%	23.9%	74.9%	**78.9%**
Case Study 8 vrijstaande	Accuracy	98.0%	**98.5%**	81.9%	**98.5%**	86.9%	83.4%	93.0%	93.0%
	Balanced accuracy	90.6%	**94.2%**	66.6%	91.4%	43.4%	28.6%	72.0%	64.2%

Discussion

- Hit-and-miss analysis
  - Wrong labels

  - Misleading definitions
  - Multi-part buildings
  - Combined subtypes

Label	Prediction	In Real Life
Galerij	MFH	MFH

Discussion

- Hit-and-miss analysis
  - Wrong labels
  - Misleading definitions

  - Multi-part buildings
  - Combined subtypes

Label	Prediction	In Real Life
Portiek	MFH	MFH (?)

Discussion

- Hit-and-miss analysis
  - Wrong labels
  - Misleading definitions
  - Multi-part buildings

  - Combined subtypes

Label	Prediction	In Real Life
Galerij	MFH	Galerij or MFH?

Discussion

- Hit-and-miss analysis
  - Wrong labels
  - Misleading definitions
  - Multi-part buildings
  - Combined subtypes

Label	Prediction	In Real Life
Portiek	MFH	MFH (or Portiekflat!)

Conclusion

- MAIN:
  - Accuracies: 61.1% - 98.5%
  - Balanced accuracies: 51.6% - 94.2%
  - Performance varies across case studies
  - Quality of results depends on input data (quality)

- FEATURES:
  - Number of adjacent buildings
  - Number of adjacent buildings of adjacent buildings
  - Number of dwellings in the building
  - Oriented bounding box width in LoD1.2

- DATA:
  - Cadastre database (BAG)
  - Semantic 3D city model (3DBAG)
  - Labelled datasets (Rijssen-Holten & EP-online)

- ML:
  - Random Forest

Conclusion

- Limitations
  - FEATURES
    - Absence of:
      - Number of floors per dwelling
      - Open porch presence
      - Galleries presence
  - DATA
    - Ground truth
      - Building type definitions
      - Preprocessing of EP-online
  - ML
    - Limited to comparison between SVM & RF

- Recommendations and future work
  - Alternative sources for missing features, for example, BAG+ (number of floors per dwelling)
  - Improve ground truth
    - Standardize classification rules
    - ML to correct ground truth labels
  - Use advanced ML algorithms
    - Neural networks

The end

Questions?

References

Figures:

- SVM: https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/
- RF: https://www.analyticsvidhya.com/blog/2021/05/bagging-25-questions-to-test-your-skills-on-random-forest-algorithm/
- Confusion matrix:
  Fawcett, T. (2006, June). An introduction to ROC analysis. Pattern Recognition Letters, 27 (8), 861–874. doi: 10.1016/j.patrec.2005.10.010