ANTICIPATING BUILDING CONSUMPTION

CITY OF SEATTLE

Maud Gesbert

SOMMAIRE

1. PROBLEM STATEMENT

Carbon-Neutral City in Terms of GHG Emissions

OBJECTIVES

- Greenhouse Gas Emissions Neutrality by 2050
- Predict the energy consumption and GHG emissions of non-residential buildings
- Evaluate the relevance of the Energy Star Score to assess the GHG emissions

1. Problem statement

2. PROVIDED DATA

Surveys conducted by city agents

RELEVÉS STRUCTURELS ET ÉNERGÉTIQUES

CONDUCTED IN 2016

On all buildings in the city

Extensive set of information for each building

Costly surveys

NumberofBuildings	NumberofFloors	PropertyGFATotal	PropertyGFAParking	PropertyGFABuilding(s)	ListOfAllPropertyUseTypes	LargestPropertyUseType	LargestPropertyUseTypeGFA
1.0	12	88434	0	88434	Hotel	Hotel	88434.0
1.0	11	103566	15064	88502	Hotel, Parking, Restaurant	Hotel	83880.0
1.0	41	956110	196718	759392	Hotel	Hotel	756493.0
1.0	10	61320	0	61320	Hotel	Hotel	61320.0
1.0	18	175580	62000	113580	Hotel, Parking, Swimming Pool	Hotel	123445.0
...	...	...	...	...	...	...	...
1.0	1	12294	0	12294	Office	Office	12294.0
1.0	1	16000	0	16000	Other - Recreation	Other - Recreation	16000.0
1.0	1	13157	0	13157	Fitness Center/Health Club/ Gym, Other - Recrea...	Other - Recreation	7583.0
1.0	1	14101	0	14101	Fitness Center/Health Club/ Gym, Food Service, ...	Other - Recreation	6601.0
1.0	1	18258	0	18258	Fitness Center/Health Club/ Gym, Food Service, ...	Other - Recreation	8271.0

STRUCTURAL ASSESSMENTS

ENERGY ASSESSMENTS

SiteEnergyUse(kBtu)	SiteEnergyUseWN(kBtu)	SteamUse(kBtu)	Electricity(kWh)	Electricity(kBtu)	NaturalGas(therms)	NaturalGas(kBtu)	DefaultData
7.226362e+06	7.456910e+06	2003882.00	1.156514e+06	3.946027e+06	12764.529300	1.276453e+06	False
8.387933e+06	8.664479e+06	0.00	9.504252e+05	3.242851e+06	51450.816410	5.145082e+06	False
7.258702e+07	7.393711e+07	21566554.00	1.451544e+07	4.952666e+07	14938.000000	1.493800e+06	False
6.794584e+06	6.946800e+06	2214446.25	8.115253e+05	2.768924e+06	18112.130860	1.811213e+06	False
1.417261e+07	1.465650e+07	0.00	1.573449e+06	5.368607e+06	88039.984380	8.803998e+06	False
...	...	...	...	...	...	...	...
8.497457e+05	9.430032e+05	0.00	1.536550e+05	5.242709e+05	3254.750244	3.254750e+05	True
9.502762e+05	1.053706e+06	0.00	1.162210e+05	3.965461e+05	5537.299805	5.537300e+05	False
5.765898e+06	6.053764e+06	0.00	5.252517e+05	1.792159e+06	39737.390630	3.973739e+06	False
7.194712e+05	7.828413e+05	0.00	1.022480e+05	3.488702e+05	3706.010010	3.706010e+05	False
1.152896e+06	1.293722e+06	0.00	1.267744e+05	4.325542e+05	7203.419922	7.203420e+05	False

Excerpt

2. Provided Data

TARGET VARIABLES

SiteEnergyUse(kBtu)	TotalGHGEmissions	ENERGYSTARScore
7.226362e+06	249.98	60.0
8.387933e+06	295.86	61.0
7.258702e+07	2089.28	43.0
6.794584e+06	286.43	56.0
1.417261e+07	505.01	75.0
...	...	...
9.320821e+05	20.94	46.0
9.502762e+05	32.17	NaN
5.765898e+06	223.54	NaN
7.194712e+05	22.11	NaN
1.152896e+06	41.27	NaN

3. DATA CLEANING AND PRELIMINARY ANALYSIS

Analysis of the dataset & Transformation of certain variables

DATA CLEANING AND PRELIMINARY ANALYSIS

1

City	State
Seattle	WA
Seattle	WA

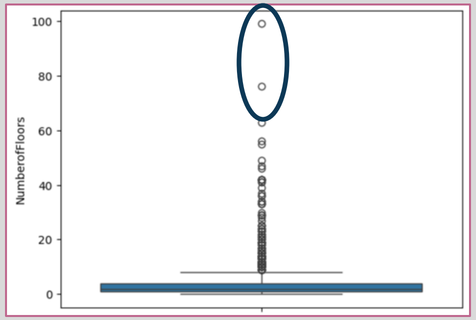
Deleting columns that are not useful

2

BuildingType	PrimaryPropertyType	PropertyName
Multifamily MR (5-9)	Mid-Rise Multifamily	Lyon Building
Multifamily MR (5-9)	Mid-Rise Multifamily	YWCA Opportunity Place
Multifamily MR (5-9)	Mid-Rise Multifamily	Wintonia
Multifamily MR (5-9)	Mid-Rise Multifamily	LAKE CITY COURT
Multifamily MR (5-9)	Mid-Rise Multifamily	Tashiro_kaplan

Deleting residential buildings

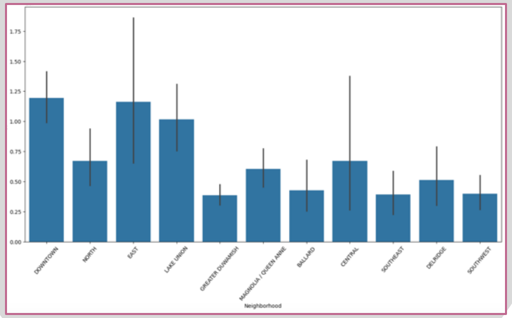
3



Deleting some outliers

DATA CLEANING AND PRELIMINARY ANALYSIS

4



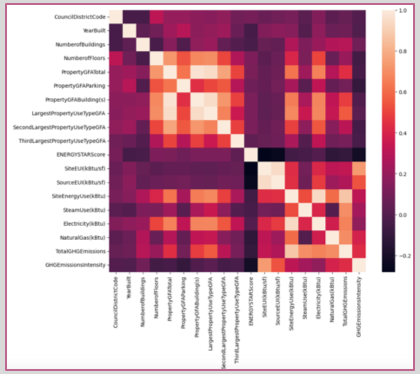
Relationship between target variables and textual variables (neighborhood)

5

LargestPropertyUseType	0
LargestPropertyUseTypeGFA	4
SecondLargestPropertyUseType	675
SecondLargestPropertyUseTypeGFA	675
ThirdLargestPropertyUseType	1152
ThirdLargestPropertyUseTypeGFA	1152
ENERGYSTARScore	530

Handling and replacement of missing values

6



Correlation analysis between variables

DATA CLEANING AND PRELIMINARY ANALYSIS

FINAL REMOVAL OF COLUMNS

Removal of "non-target" columns and those unavailable at the time of building construction

```
Index(['index', 'BuildingType', 'PrimaryPropertyType', 'CouncilDistrictCode',
       'Neighborhood', 'YearBuilt', 'NumberofBuildings', 'NumberofFloors',
       'PropertyGFATotal', 'PropertyGFAParking', 'PropertyGFABuilding(s)',
       'LargestPropertyUseType', 'LargestPropertyUseTypeGFA',
       'SecondLargestPropertyUseType', 'SecondLargestPropertyUseTypeGFA',
       'ThirdLargestPropertyUseType', 'ThirdLargestPropertyUseTypeGFA',
       'ENERGYSTARScore', 'SiteEUI(kBtu/sf)', 'SourceEUI(kBtu/sf)',
       'SiteEnergyUse(kBtu)', 'SteamUse(kBtu)', 'Electricity(kBtu)',
       'NaturalGas(kBtu)', 'TotalGHGEmissions', 'GHGEmissionsIntensity'],
```

REMAINING COLUMNS

```
Data columns (total 20 columns):
 #   Column
---  ------
 0   index
 1   BuildingType
 2   PrimaryPropertyType
 3   CouncilDistrictCode
 4   Neighborhood
 5   YearBuilt
 6   NumberofBuildings
 7   NumberofFloors
 8   PropertyGFATotal
 9   PropertyGFAParking
 10  PropertyGFABuilding(s)
 11  LargestPropertyUseType
 12  LargestPropertyUseTypeGFA
 13  SecondLargestPropertyUseType
 14  SecondLargestPropertyUseTypeGFA
 15  ThirdLargestPropertyUseType
 16  ThirdLargestPropertyUseTypeGFA
 17  ENERGYSTARScore
 18  SiteEnergyUse(kBtu)
 19  TotalGHGEmissions
```

CATEGORICAL VARIABLE TRANSFORMATION

LargestPropertyUseType	SecondLargestPropertyUseType	ThirdLargestPropertyUseType
Convention Center	Parking	Financial Office
Office	Laboratory	Non-Refrigerated Warehouse
Hotel	Parking	Parking
Office	Parking	Parking
Hospital (General Medical & Surgical)	Parking	Other
Retail Store	Other	Financial Office
Office	Parking	Retail Store
Office	Parking	Office
Parking	Multifamily Housing	Medical Office
Parking	Other - Entertainment/Public Assembly	Parking
Retail Store	Office	Office
Office	Parking	Restaurant
Multifamily Housing	Hotel	Office
Parking	Retail Store	Hotel
Office	Parking	Multifamily Housing
Medical Office	Parking	Multifamily Housing
Office	Retail Store	Data Center
Non-Refrigerated Warehouse	Refrigerated Warehouse	Other/Specialty Hospital
Office	Parking	Other
Medical Office	Parking	Other
Hospital (General Medical & Surgical)	Parking	Restaurant
Office	Parking	Restaurant
Office	Parking	Other
Office	Parking	Multifamily Housing
Data Center	Office	Multifamily Housing
Office	Parking	Fitness Center/Health Club/Gym
Office	Parking	Office
Office	Parking	

Set of possible values

7.5.1 LargestPropertyUseType

```
buildings['LargestPropertyUseType'].unique()
```

```
array(['Hotel', 'Police Station', 'Other - Entertainment/Public Assembly',
       'Library', 'Fitness Center/Health Club/Gym', 'Social/Meeting Hall',
       'Courthouse', 'Other', 'College/University',
       'Automobile Dealership', 'Office', 'Self-Storage Facility',
       'Non-Refrigerated Warehouse', 'K-12 School', 'Other - Mall',
       'Medical Office', 'Retail Store',
       'Hospital (General Medical & Surgical)', 'Museum',
       'Repair Services (Vehicle, Shoe, Locksmith, etc)',
       'Other - Lodging/Residential', 'Other/Specialty Hospital',
       'Financial Office', 'Distribution Center', 'Parking',
       'Multifamily Housing', 'Worship Facility', 'Restaurant',
       'Data Center', 'Laboratory', 'Supermarket/Grocery Store',
       'Urgent Care/Clinic/Other Outpatient', nan, 'Other - Services',
       'Strip Mall', 'Wholesale Club/Supercenter',
       'Refrigerated Warehouse', 'Manufacturing/Industrial Plant',
       'Other - Recreation', 'Lifestyle Center',
       'Other - Public Services', 'Fire Station', 'Performing Arts',
       'Residential Care Facility', 'Bank Branch', 'Other - Education',
       'Other - Restaurant/Bar', 'Food Service', 'Adult Education',
       'Other - Utility', 'Movie Theater',
       'Personal Services (Health/Beauty, Dry Cleaning, etc)',
       'Residence Hall/Dormitory', 'Pre-school/Daycare',
       'Prison/Incarceration'], dtype=object)
```

CATEGORICAL VARIABLE TRANSFORMATION

Reduction of possibilities

```
Largest_value_1 = 'Office'
Largest_value_2 = 'Hospital'
Largest_value_3 = 'Warehouse'
Largest_value_4 = 'School'
Largest_value_5 = 'Repair and Public Services'
Largest_value_6 = 'Food/Drink Services'
Largest_value_7 = 'Retail/Mall'
Largest_value_8 = 'Recreational Venues'

# Creating a dictionary to be able to use the replace methods to handle strings with parenthesis.
replacement_mapping = {
    'Medical Office': Largest_value_1,
    'Office': Largest_value_1,
    'Financial Office': Largest_value_1,
    'Hospital (General Medical & Surgical)': Largest_value_2,
    'Other/Specialty Hospital': Largest_value_2,
    'Urgent Care/Clinic/Other Outpatient': Largest_value_2,
    'Non-Refrigerated Warehouse': Largest_value_3,
    'Self-Storage Facility': Largest_value_3,
    'Distribution Center': Largest_value_3,
    'Refrigerated Warehouse': Largest_value_3,
    'College/University': Largest_value_4,
    'K-12 School': Largest_value_4,
    'Other - Education': Largest_value_4,
    'Adult Education': Largest_value_4,
    'Pre-school/Daycare': Largest_value_4,
    'Repair Services (Vehicle, Shoe, Locksmith, etc)': Largest_value_5,
    'Other - Services': Largest_value_5,
    'Other - Public Services': Largest_value_5,
    'Personal Services (Health/Beauty, Dry Cleaning, etc)': Largest_value_5,
    'Restaurant': Largest_value_6,
    'Other - Restaurant/Bar': Largest_value_6,
    'Food Service': Largest_value_6,
    'Supermarket/Grocery Store': Largest_value_6,
    'Other - Mall' : Largest_value_7,
    'Strip Mall' : Largest_value_7,
    'Retail Store' : Largest_value_7,
    'Wholesale Club/Supercenter' : Largest_value_7,
    'Other - Entertainment/Public Assembly' : Largest_value_8,
    'Other - Recreation' : Largest_value_8,
    'Social/Meeting Hall' : Largest_value_8,
    'Movie Theater' : Largest_value_8
}

# Replacing the values
buildings['LargestPropertyUseType'] = buildings['LargestPropertyUseType'].replace(replacement_mapping)
```

```
buildings['LargestPropertyUseType'].unique()

array(['Hotel', 'Police Station', 'Recreational Venues', 'Library',
       'Fitness Center/Health Club/Gym', 'Courthouse', 'Other', 'School',
       'Automobile Dealership', 'Office', 'Warehouse', 'Retail/Mall',
       'Hospital', 'Museum', 'Repair and Public Services',
       'Other - Lodging/Residential', 'Parking', 'Multifamily Housing',
       'Worship Facility', 'Food/Drink Services', 'Data Center',
       'Laboratory', nan, 'Manufacturing/Industrial Plant',
       'Lifestyle Center', 'Fire Station', 'Performing Arts',
       'Residential Care Facility', 'Bank Branch', 'Other - Utility',
       'Residence Hall/Dormitory', 'Prison/Incarceration'], dtype=object)
```

CATEGORICAL VARIABLE TRANSFORMATION
ENCODING

Removal of the original column

Addition of new columns

0 and 1

LargestPropertyUseType_Hotel	LargestPropertyUseType_Laboratory	LargestPropertyUseType_Library	LargestPropertyUseType_Lifestyle Center	LargestPropertyUseType_Manufacturing/ Industrial Plant
1.0	0.0	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0
...	...	...	...	...
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0

ANALYSIS METRICS

REMOVAL OF UNNECESSARY COLUMNS AND OUTLIERS

Initially: 3,376 rows × 46 columns

Residential: 1,668 rows × 46 columns

Outliers + Unnecessary columns:

1,477 rows × 20 columns

56% of rows & 57% of columns removed

ENCODING

Value transformation

1,477 rows × 112 columns

4. MODELING

Implementation of machine learning models

Models selection

Models

- Continuous values → Regression

- Linear Regression
- Random Forest
- Gradient Boosting
- Support Vector Regression

Metrics

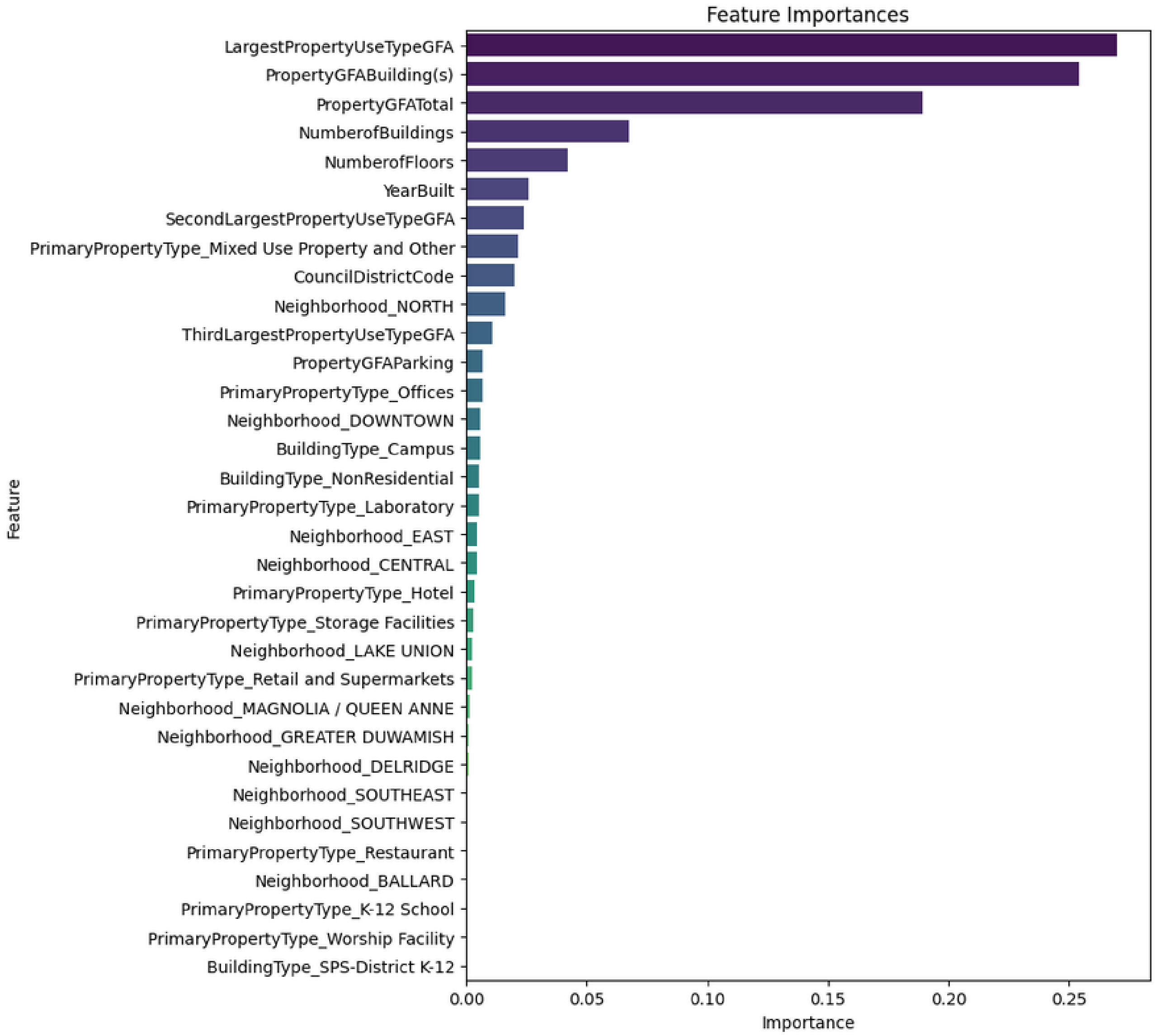
- R^2 (Coefficient of determination)
- Mean Squared Error
- Execution Time
- Number of variables used

PROCESS

- All variables
- Selection of relevant variables
- Selection of model hyperparameters (number of trees, tree depth, etc.)
- Results analysis and model execution time

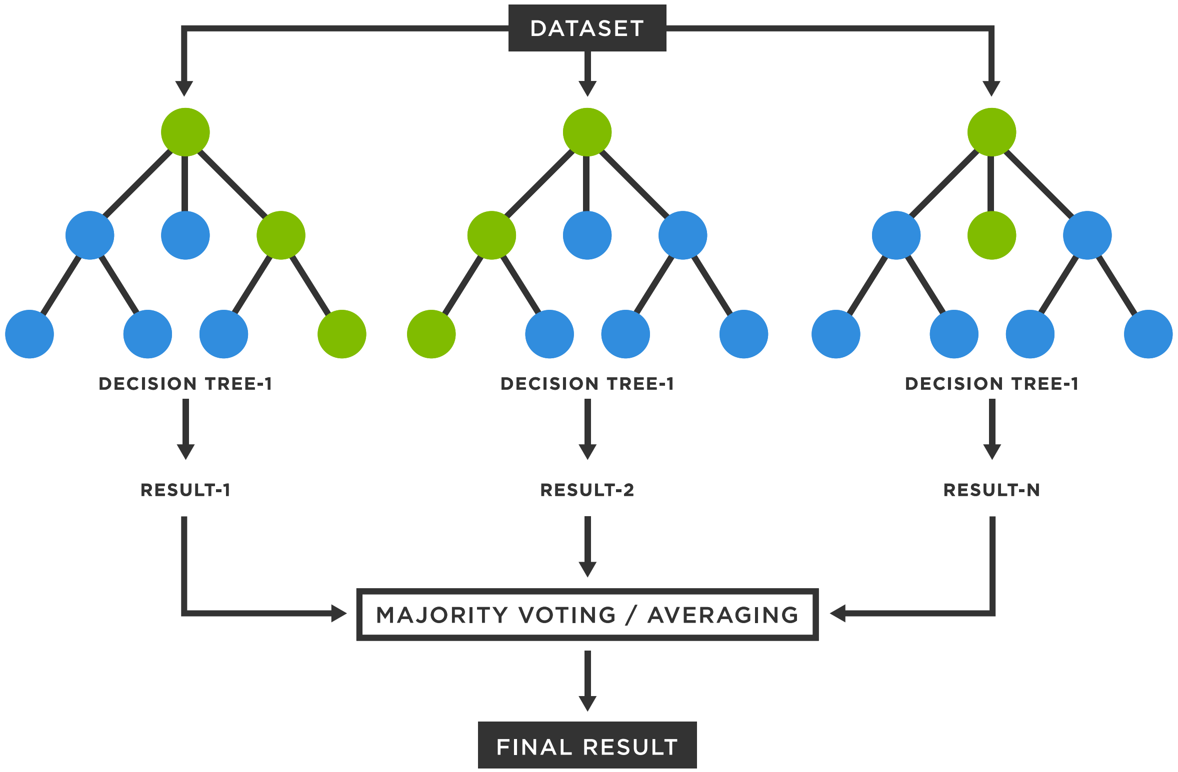
RANDOM FOREST

- Relevant variable selection



4. Modeling

RANDOM FOREST
SELECTION OF MODEL HYPERPARAMETERS



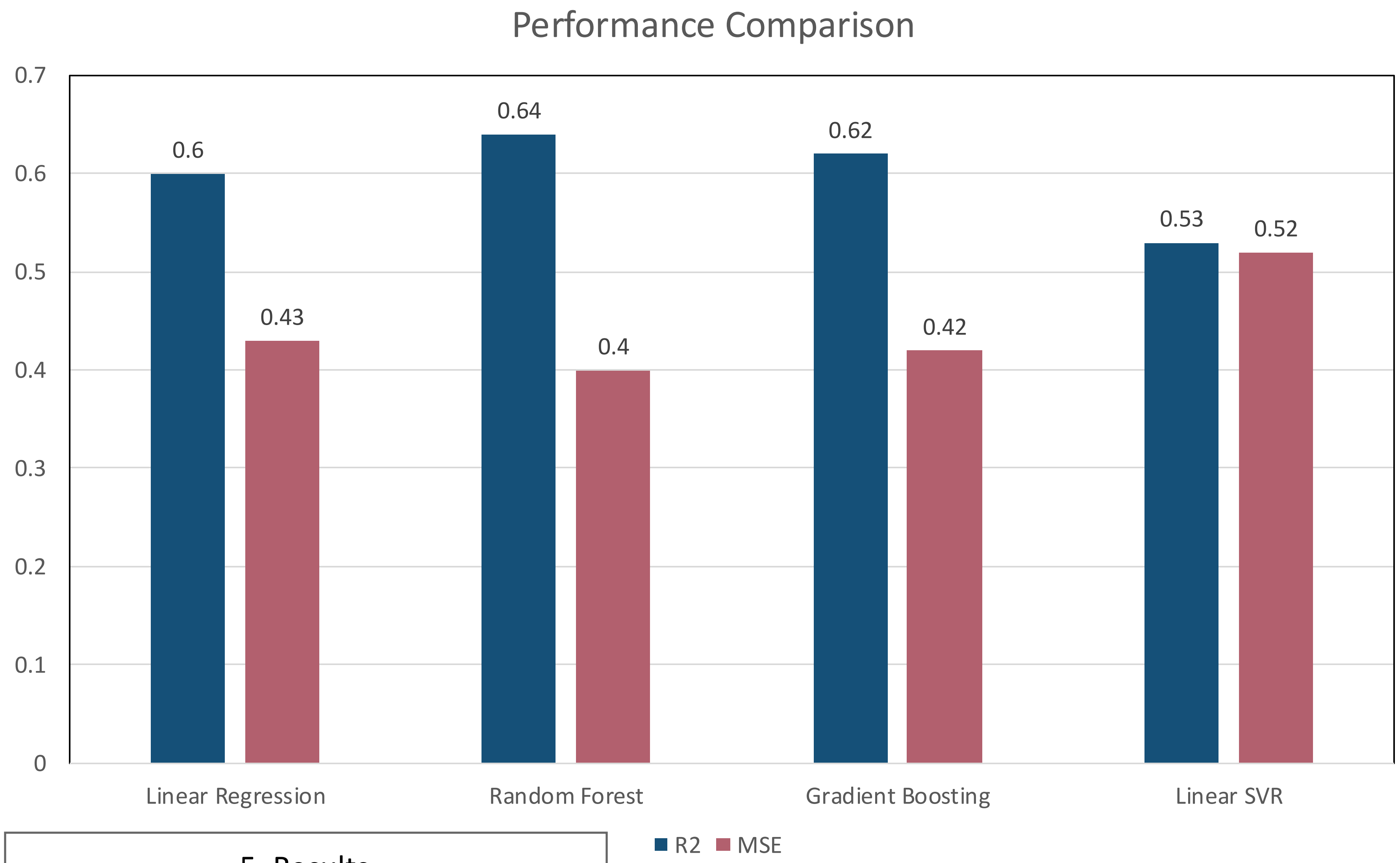
* Source : MétéoSuisse-Blog | 30 octobre 2022

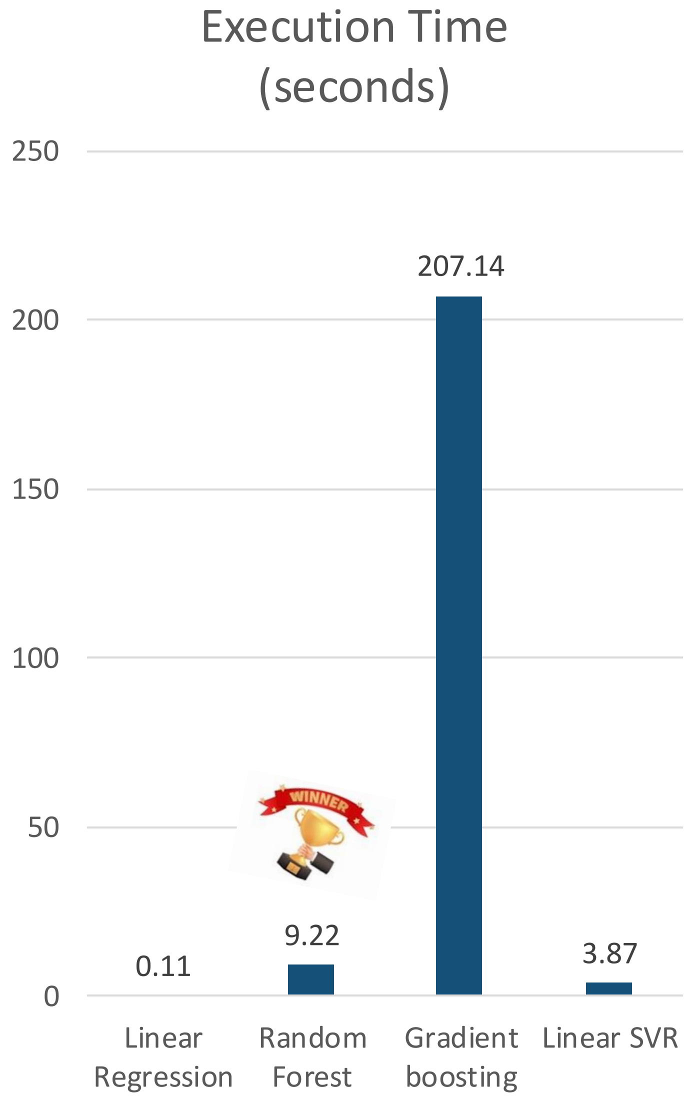
- Number of trees
- Maximum tree depth
- Minimum number of samples required to split an internal node
- Minimum number of samples required to be at a leaf node
- Threshold: Selection of variables with an importance score above a certain value

5. RESULTS

Models comparison

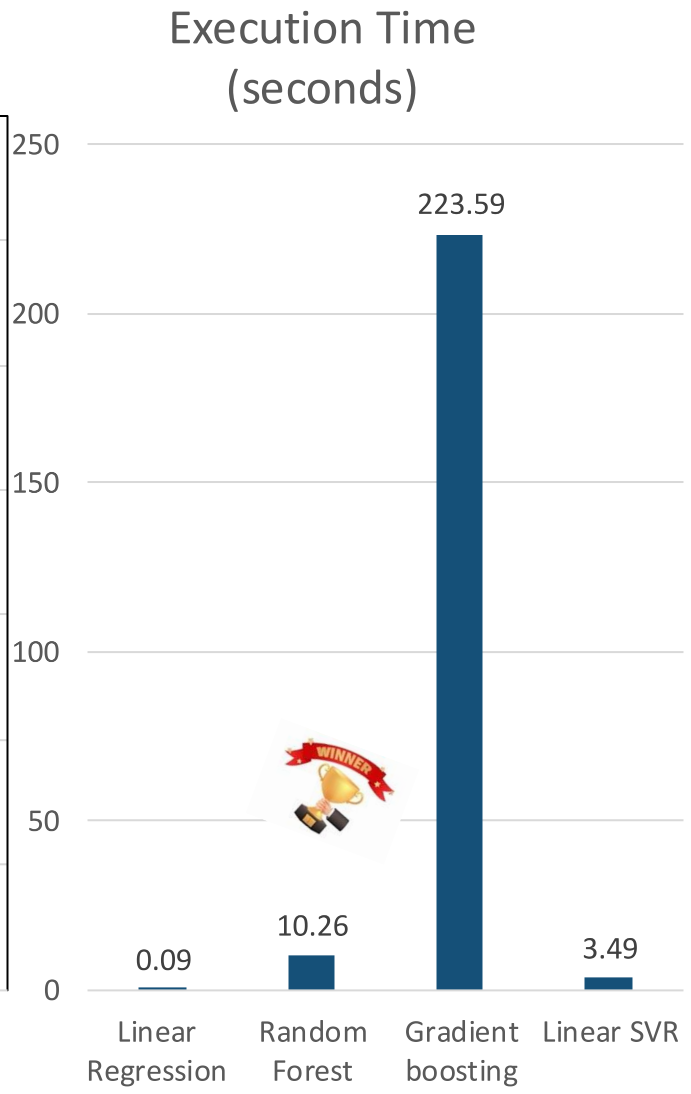
ALL MODELS – CONSUMPTION





5. Results

ALL MODELS – EMISSIONS

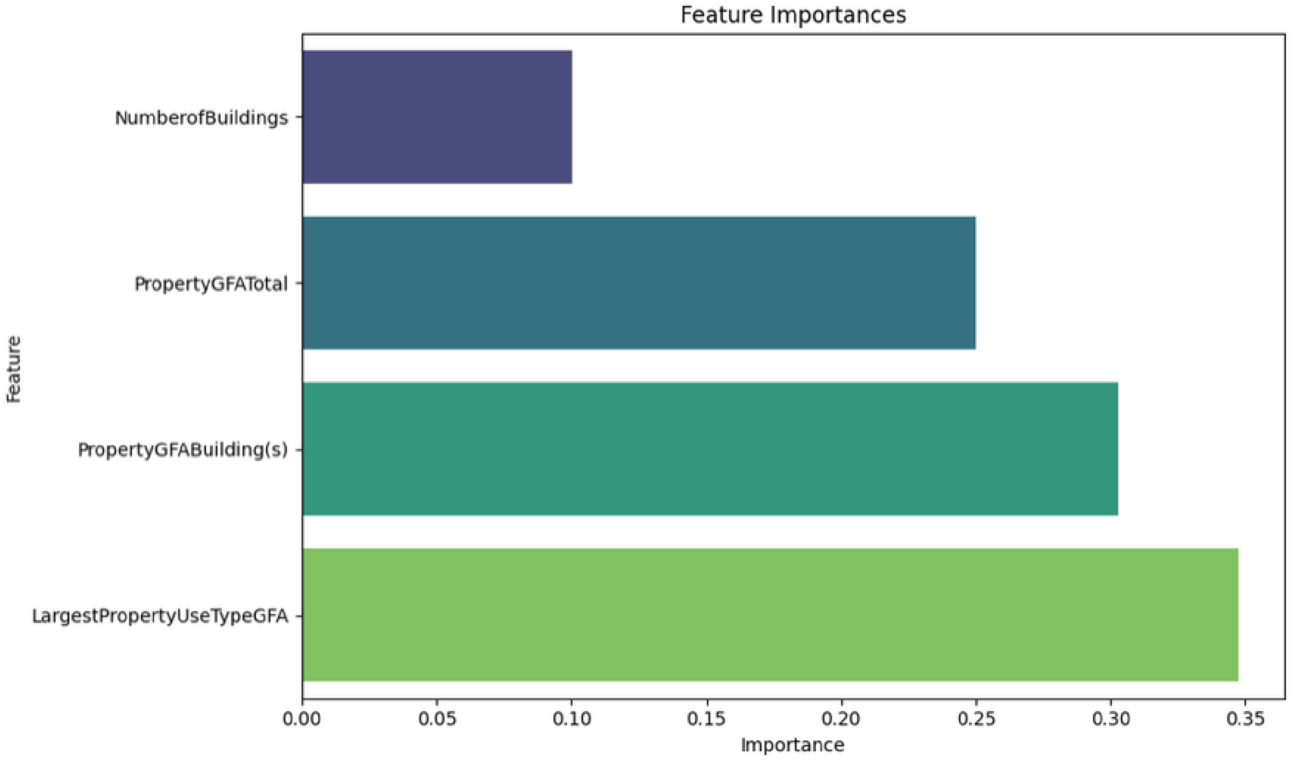


5. Results

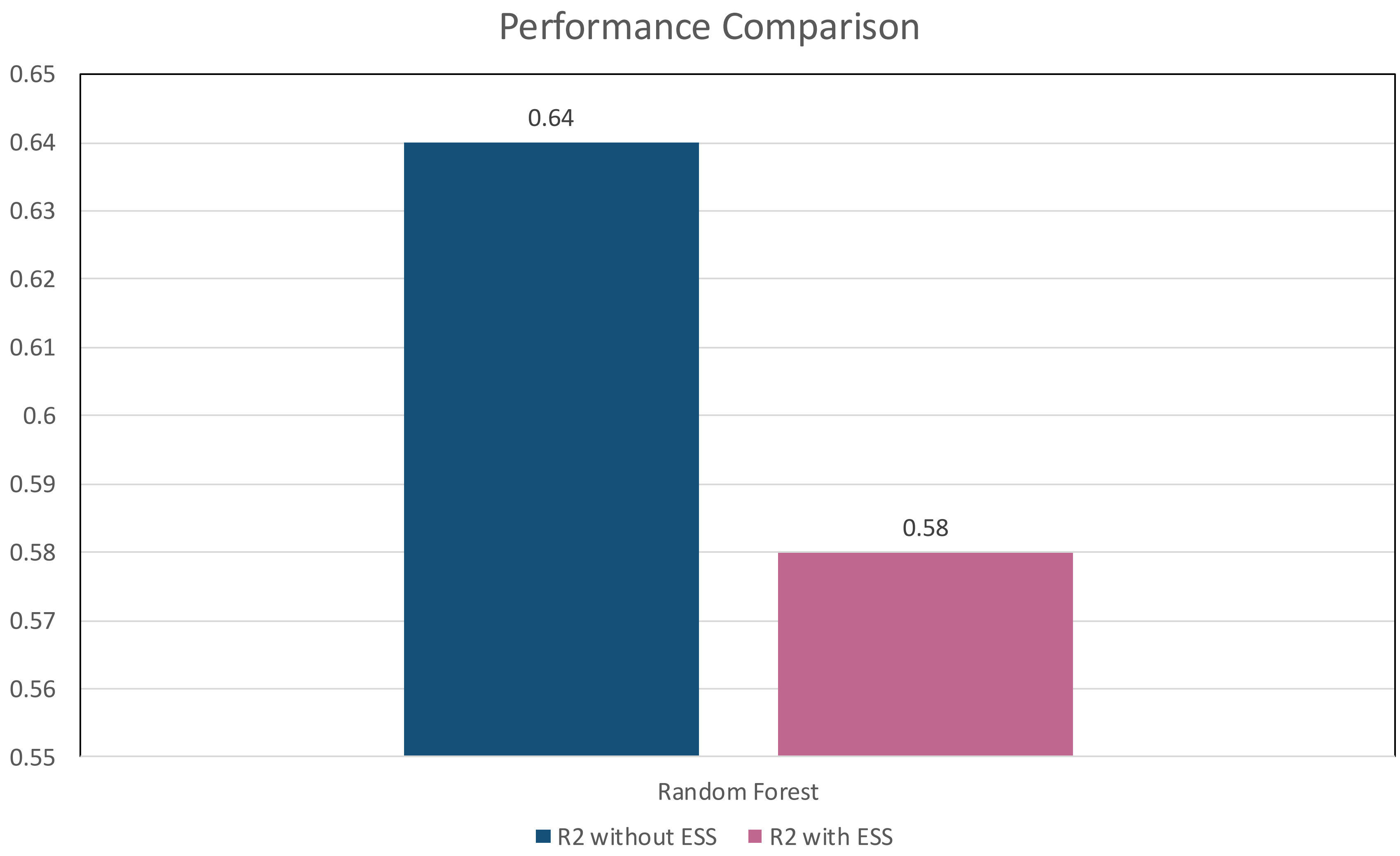
RANDOM FOREST– HYPERPARAMETERS

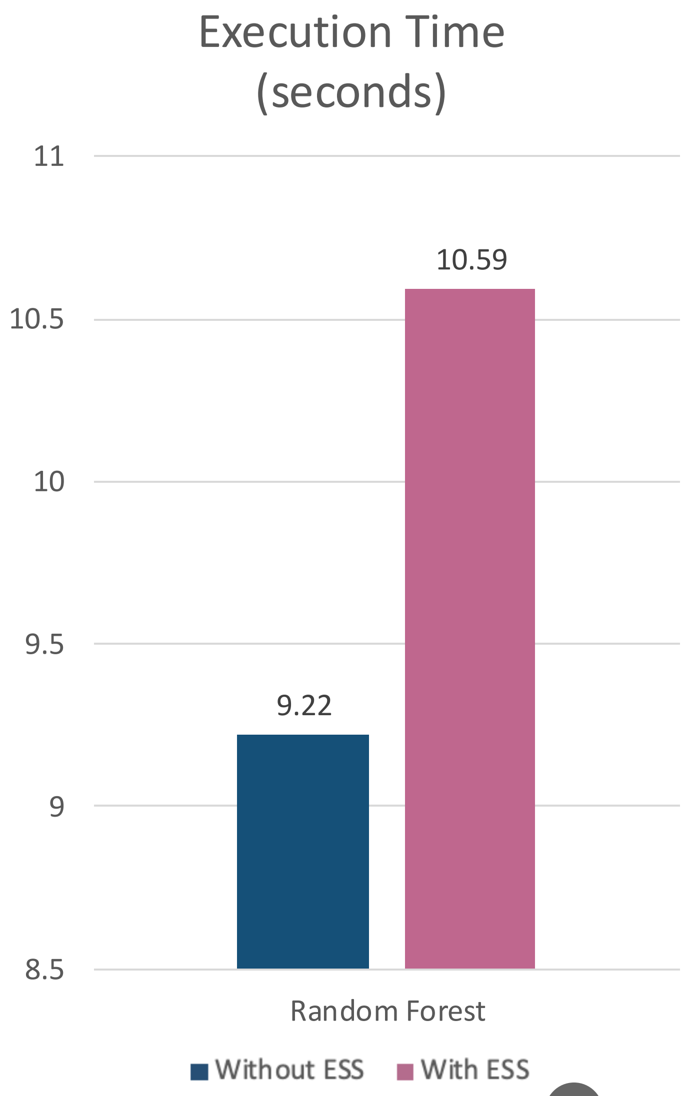
- Number of trees – 500
- Maximum tree depth – None
- Minimum number of samples required to split an internal node – 2
- Minimum number of samples required to be at a leaf node – 1
- Threshold: Selection of variables with an importance score above a certain value – Threshold set at 0.05

```
Feature Importances:
NumberofBuildings: 0.1001
PropertyGFATotal: 0.2502
PropertyGFABuilding(s): 0.3026
LargestPropertyUseTypeGFA: 0.3471
```



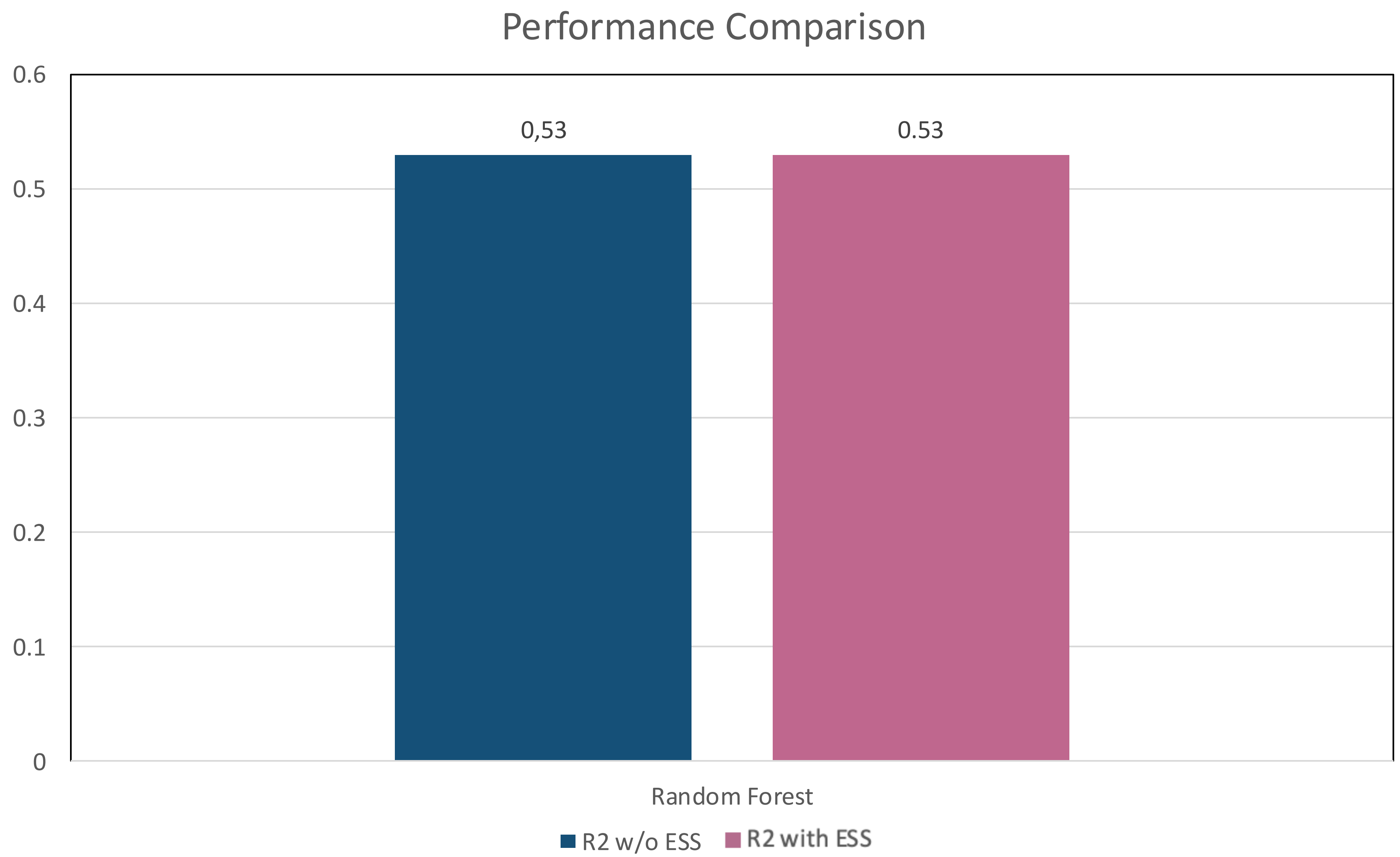
ENERGY STAR SCORE - CONSUMPTION

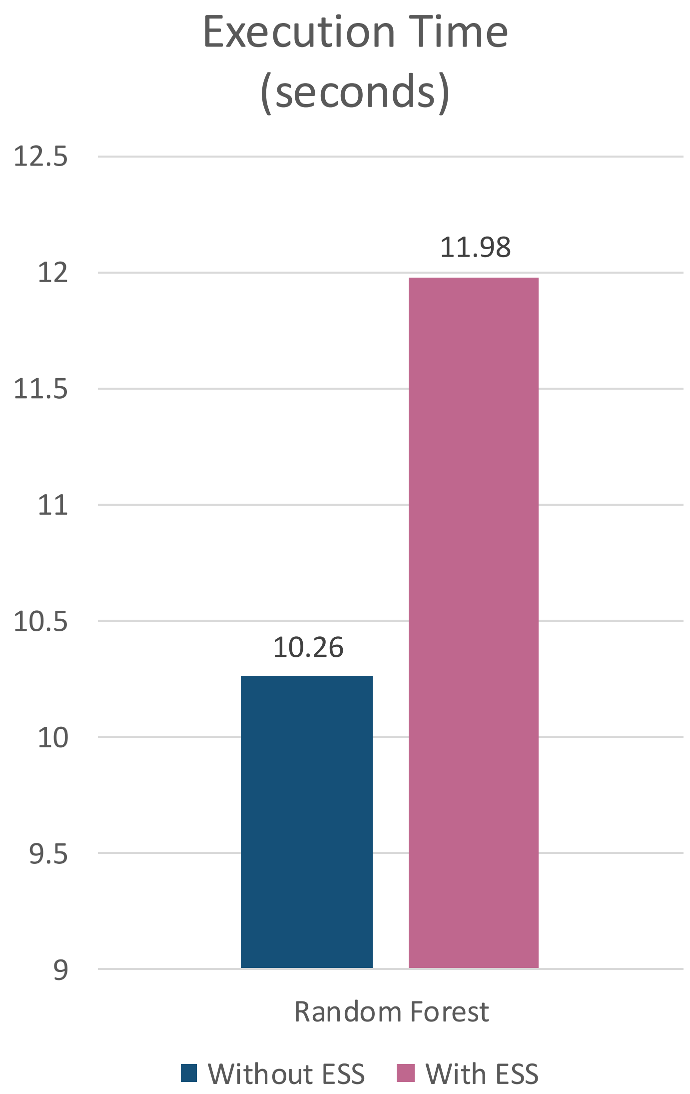




5. Results

ENERGY STAR SCORE - ÉMISSIONS





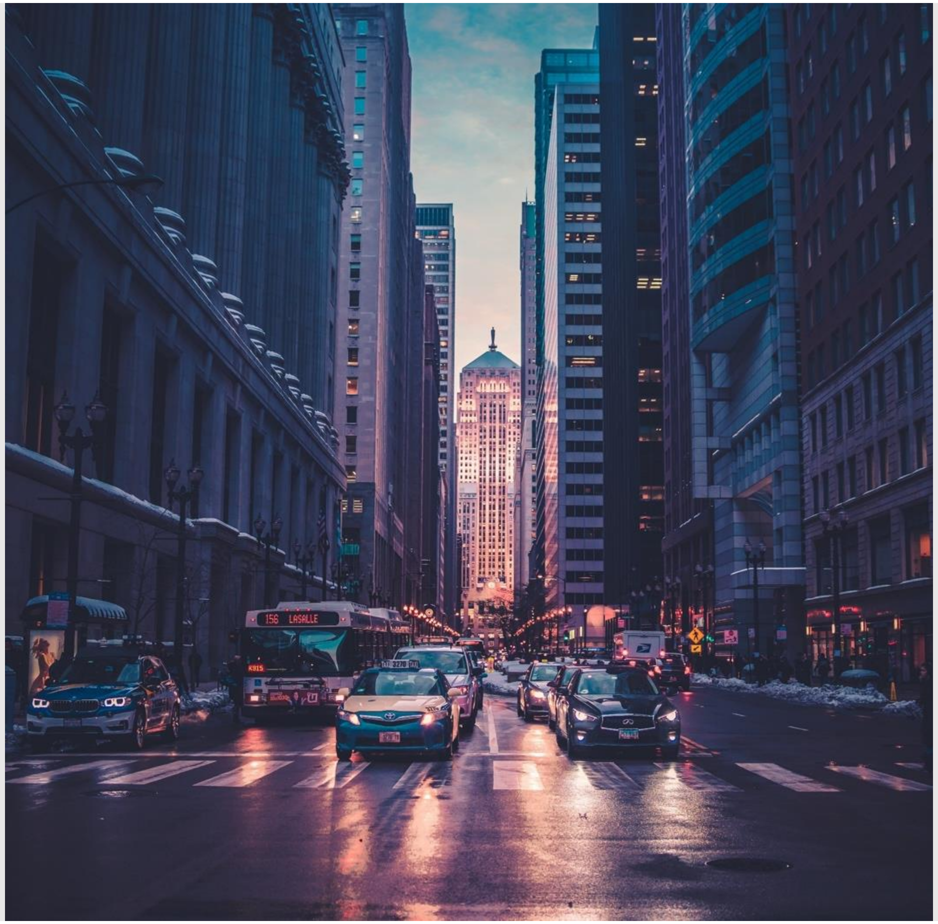
5. Results

ENERGY STAR SCORE

CONSUMPTION

- Addition of the Energy Star Score to the input variables

Degrades the quality of the predictions



EMISSIONS

- Addition of the Energy Star Score to the input variables

Has no impact on prediction performance

6. CONCLUSION

CONCLUSION

- Inconclusive results – R^2 does not exceed 0.64 (At least 0.7 is usually required to deploy a model in production)
  - Limited number of buildings: (Non-residential buildings represent 1,668 rows — a small sample)

- ML improvement opportunities:
  - Improved feature engineering
  - More thorough hyperparameter tuning (Note: Some hyperparameter searches take a long time to complete)
  - Testing other types of regression models

THANK YOU