Python Data Scinece

mentor:
Vladimir Kovačević

candidate:
Snežana Zurovac

Title: Machine Learning model -Regression

- **GOAL**:

  BUILD MACHINE LEARNING MODEL THAT CAN BE USED IN REAL TIME APPLICATION

- **WHO CAN USE IT**:

  companies that rent bikes

- **HOW MODEL CAN BE USED**:

  To predict demand of bike rental in specific period of time, and company can fit organization, investment plans, to be prepared for customer request

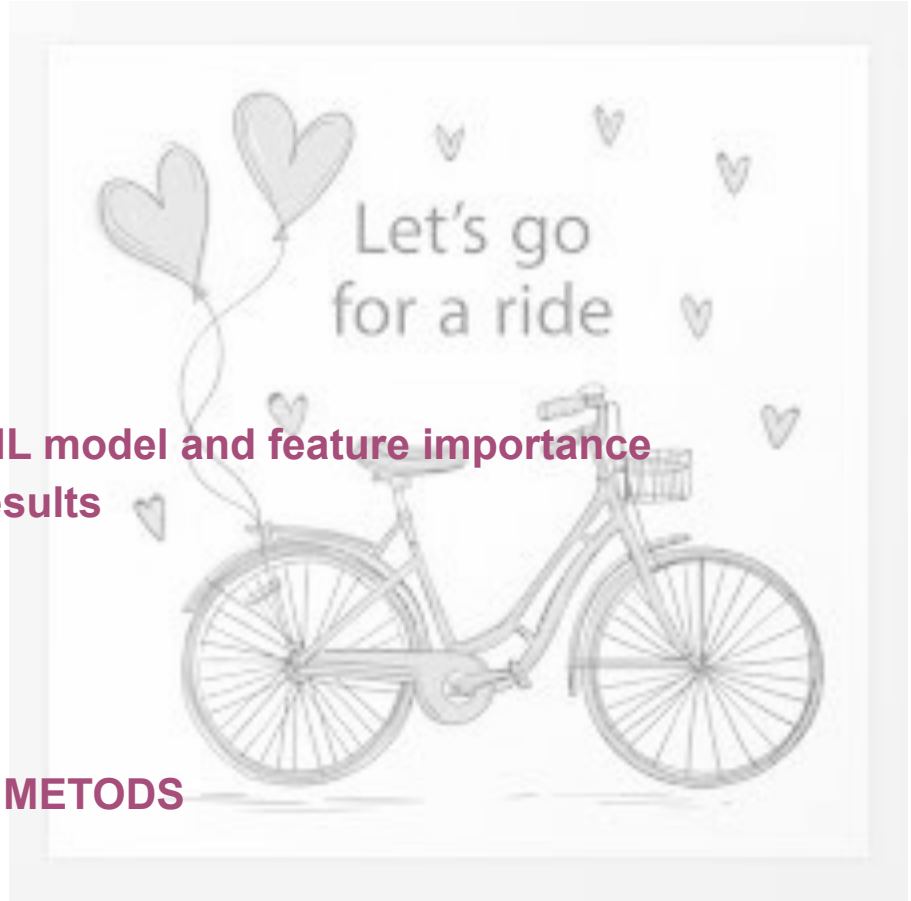
- **DATA SET**: Bike Sharing Dataset

  Bike sharing systems are new generation of traditional bike rentals where whole process from membership, rental and return back has become automatic

BUILDING MACHINE LEARNING MODEL CONSISTS NEXT STEPS:

I-BUILDING ML_MODEL

1.Importing dataset and necessary libraries
2.Exploratory Data Analyze
3.Preprocessing data set
4.Building model
5.Estimating and analyze all parametar of ML model and feature importance
6.Test model and analyze of the obtained results



II-IMPORVING ML_MODEL

1.Improving Ml_model by using pipeline
2.Impoving Ml_model by using ENSEMBLE METODS
(Voting Regressor and Stacking metod)

I-BUILDING ML_MODEL

Exploratory Data Analyze- gives the answers to the next question of obesereved bike renting data set:

1.When company can anticipate increase of bike renting?

2.What happens true the different months?

3.What is the trend of bike renting true the years?

4.How bike renting depends on the seasons ?

Model Random Forest Regressor gives the best result

Diagram of MEAN APSOLUT ERROR according to the model

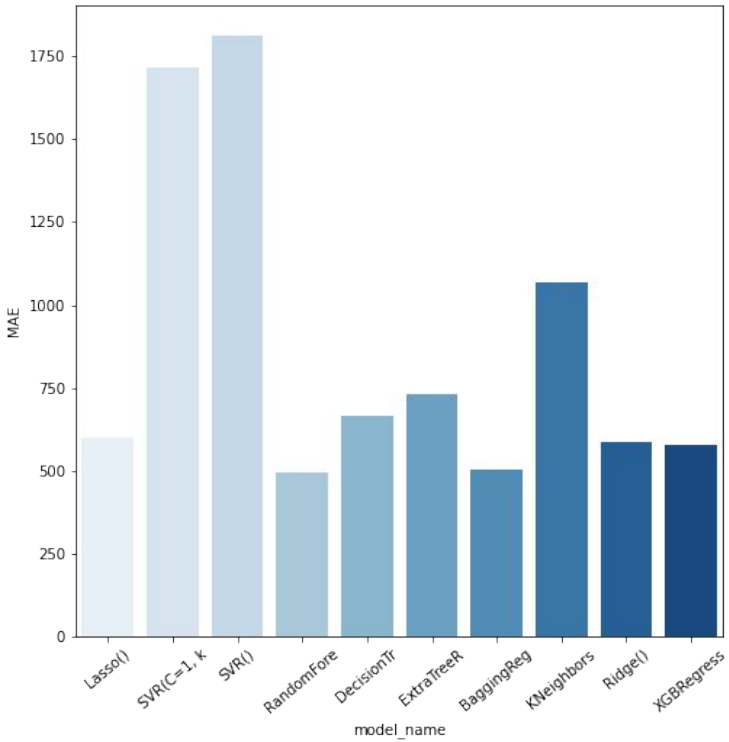
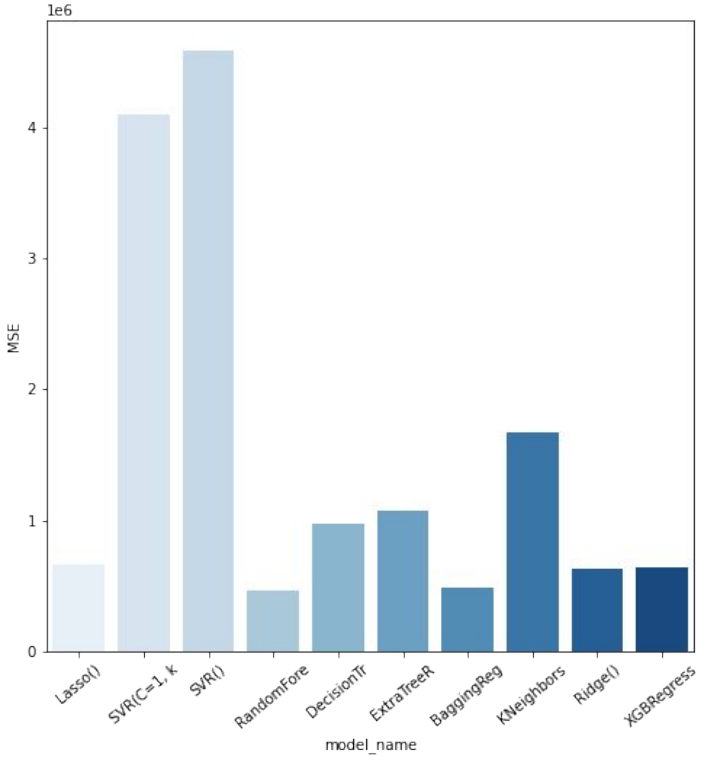
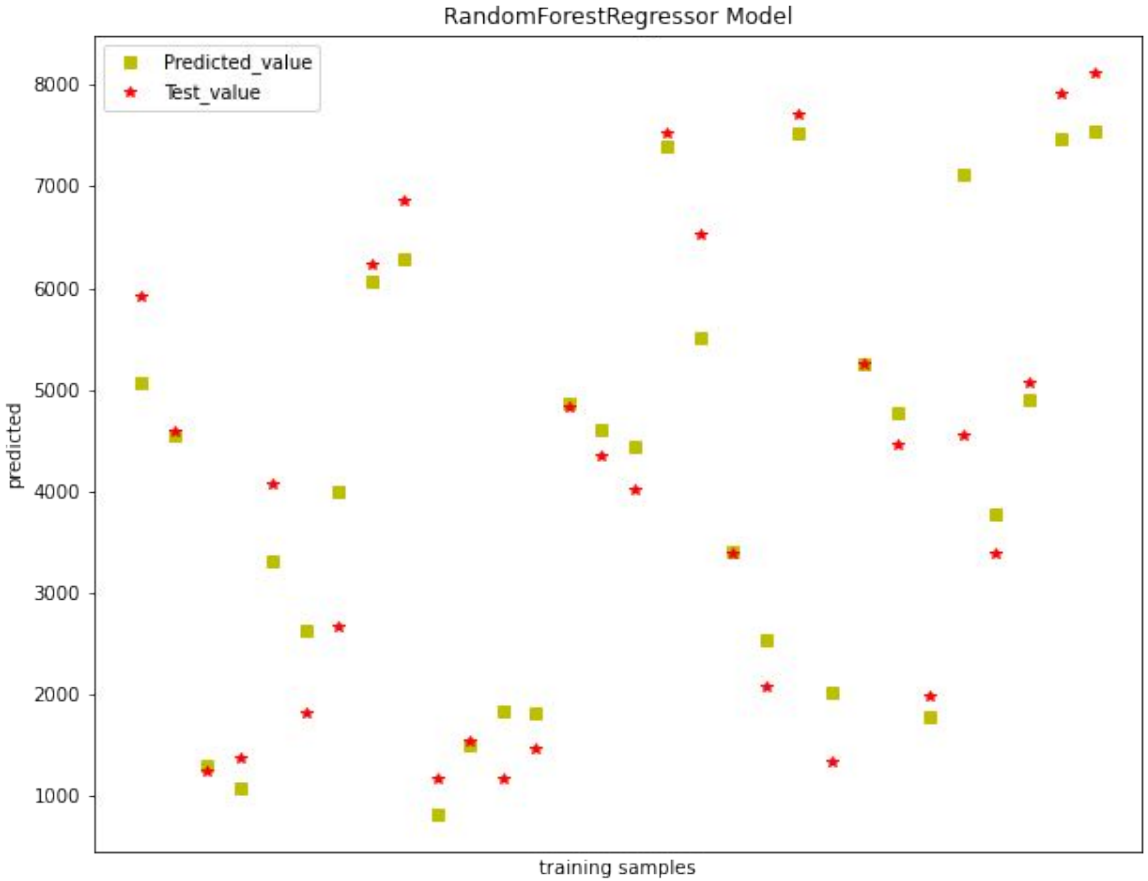


Diagram of Mean SQUERED ERROR according to the model



Plot results of prediction for first 30 row of test data



MODEL IMPROVEMENT - PIPELINE

model	MSE	MAE	R2
RandomForestReg	455360.65	443.59	0.88644

Pipeline is very elegant method for data prepocessing and applying model. Hance, it gives good results it should be always used always when it's possible.

MODEL IMPROVEMENT - ENSEMBLE METHODS

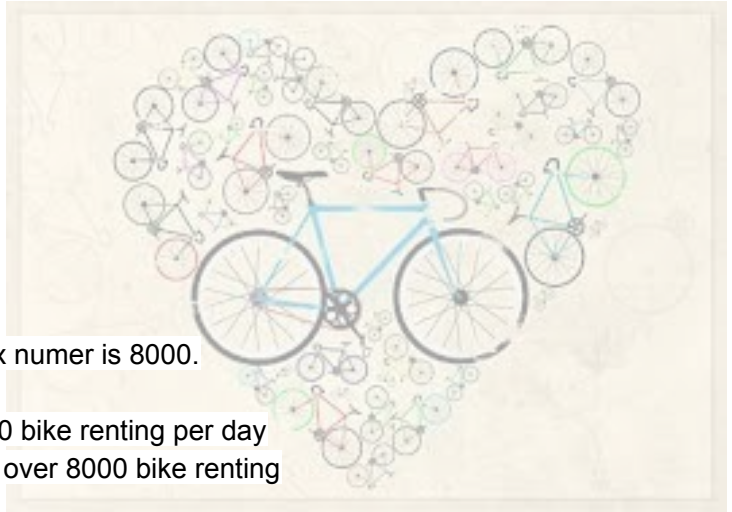
1.Ensemble method_ Stacking method

2.Ensemble method _Voting Regressor

	MSE	MAE	R2	modeL
0	511.247545	507977.253133	0.890594	GradientBoostingRegressor
1	516.384818	497769.136897	0.892792	RandomForestRegressor
2	599.790909	660123.245455	0.857825	LinearRegression
3	495.014334	457942.981055	0.901370	Ensemble

Ensemble method _Voting Regressor gives the best results

Conclusion:



1.When company can anticipate increasing of bike renting?

Answer:seqnificunt increase of count of bike renting is in seasons:

a)fall and summer with over 6000 bike renting

b)winter is up to 6000

c)spring is below 6000

2.What happens true the different months?

Increase of bike renting sarts in may with avrage 4000 thousend bike renting and max numer is 8000.

Increasing continues in next four mounth:

1.Jun,July and August with avrage 5000 numbers and maximum number reaches 8000 bike renting per day

2.In September number of counting avrage is about 5500,and reches maximum value over 8000 bike renting

3.In October nubmer of counting also has same trend as in September

4.In November and December number is decreased, below 5000

5. January and February are the months when number of bike renting has been halved,and in those months is below 3000

3.What is the trend of bike renting true the years?

Number of bike renting is significantly increased in two years,sach that it was doubled.In 2011 it was 3500 renting,and in 2012 it reached 7000 renting.

4.How bike renting depends on the seasons ?

The highest value is when the weather site is:

1: Clear, Few clouds, Partly cloudy

2:It decrease if weather is : Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

3.It is lowest when weather is : Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds

4.It does not exist if weather is : Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

How model can be used for bussines improvment purposes?

All, given facts about bike renting occurrence,can be used for:

1.Making plans for further investition

2.Planing time of bike reparing

3.Planing of maximum and minimum number of bikes that have to be prepared for renting

4.Also, analize should give new request for information which are important for prediction an business imporvment

The entire project can be seen in Jupyter Note Book on the link:

https://github.com/snezanazurovac/My-DataScince-Centar.git

Thank you for your attention !!!