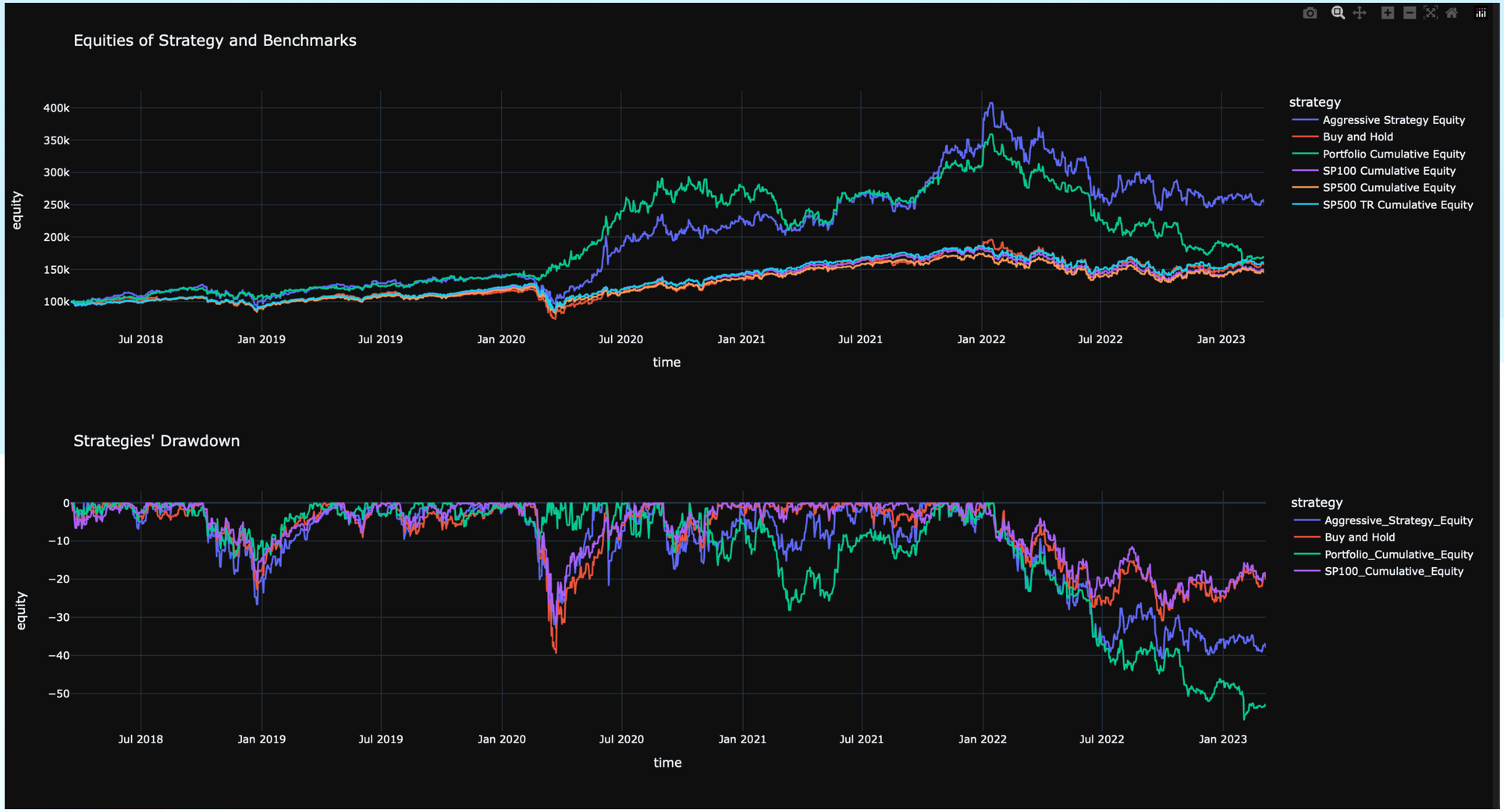
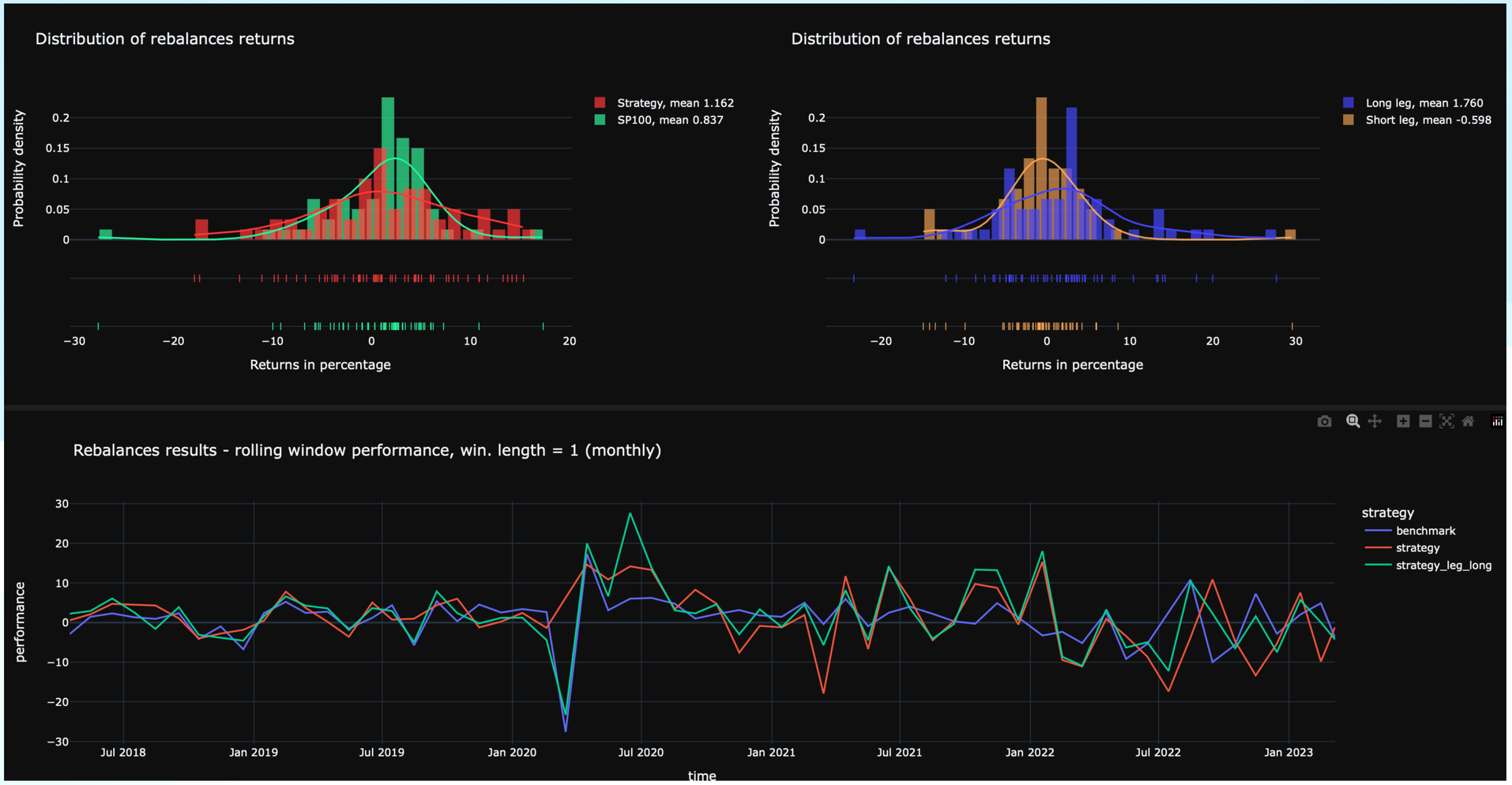
Portfolio of projects

Vadym Hladyuk

Data visualisation

- Application for visualisation of portfolio performance and characteristics
- Dash python - Data Visualization Interfaces
- Interactive web application
- Visualization of portfolio evolution, portfolio metrics and other portfolio information
- In presentation it is only pictures but application in web is interactive


Equities of Strategy and Benchmarks
equity
400k
350k
300k
250k
200k
150k
100k
Jul 2018
Jan 2019
Jul 2019
Jan 2020
Jul 2020
Jan 2021
Jul 2021
Jan 2022
Jul 2022
Jan 2023
time
strategy
Aggressive Strategy Equity
Buy and Hold
Portfolio Cumulative Equity
SP100 Cumulative Equity
SP500 Cumulative Equity
SP500 TR Cumulative Equity
Strategies' Drawdown
equity
0
−10
−20
−30
−40
−50
Jul 2018
Jan 2019
Jul 2019
Jan 2020
Jul 2020
Jan 2021
Jul 2021
Jan 2022
Jul 2022
Jan 2023
time
strategy
Aggressive_Strategy_Equity
Buy and Hold
Portfolio_Cumulative_Equity
SP100_Cumulative_Equity


Distribution of rebalances returns
Probability density
0.2
0.15
0.1
0.05
0
Strategy, mean 1.162
SP100, mean 0.837
−30
−20
−10
0
10
20
Returns in percentage
Distribution of rebalances returns
Probability density
0.2
0.15
0.1
0.05
0
Long leg, mean 1.760
Short leg, mean -0.598
−20
−10
0
10
20
30
Returns in percentage
Rebalances results - rolling window performance, win. length = 1 (monthly)
performance
30
20
10
0
−10
−20
−30
Jul 2018
Jan 2019
Jul 2019
Jan 2020
Jul 2020
Jan 2021
Jul 2021
Jan 2022
Jul 2022
Jan 2023
time
strategy
benchmark
strategy
strategy_leg_long

Data analysis in tennis

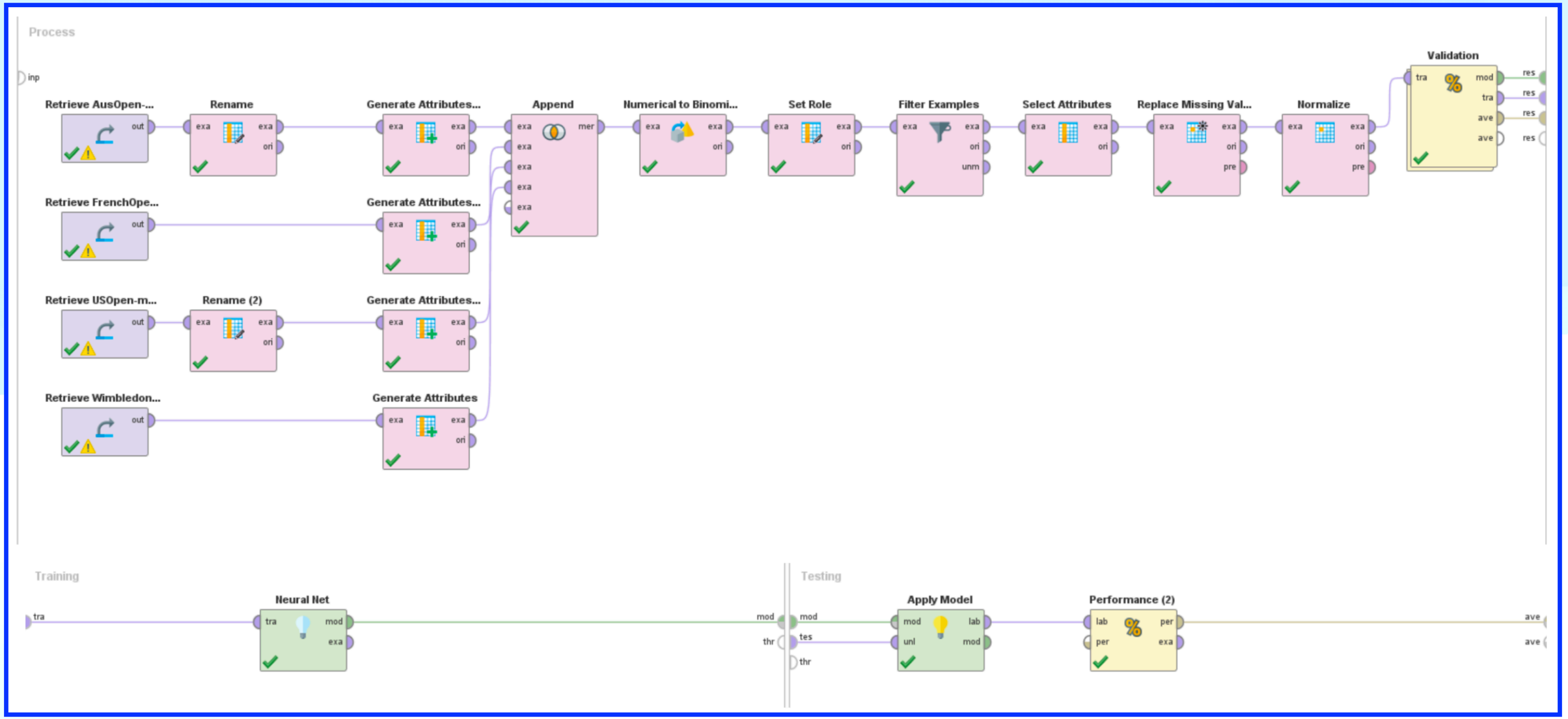
Predicting the last set and grouping players by serve

- School project in Rapid Miner
- No-code platform
- Tennis tournament dataset (from serve data to number of games won)

Data analysis in tennis

Prediction result of the last set (only 5th set)

- Converting input to correct data types and filling in missing values
- Filtering data
- Predicting the result of the fifth set
- Two models: Neural Network and SVM


Process
inp
Retrieve AusOpen-...
Rename
Generate Attributes...
Append
Numerical to Binomi...
Set Role
Filter Examples
Select Attributes
Replace Missing Val...
Normalize
Validation
Retrieve FrenchOpe...
Generate Attributes...
Retrieve USOpen-m...
Rename (2)
Generate Attributes...
Retrieve Wimbledon...
Generate Attributes
Training
Neural Net
Testing
Apply Model
Performance (2)

Data analysis in tennis

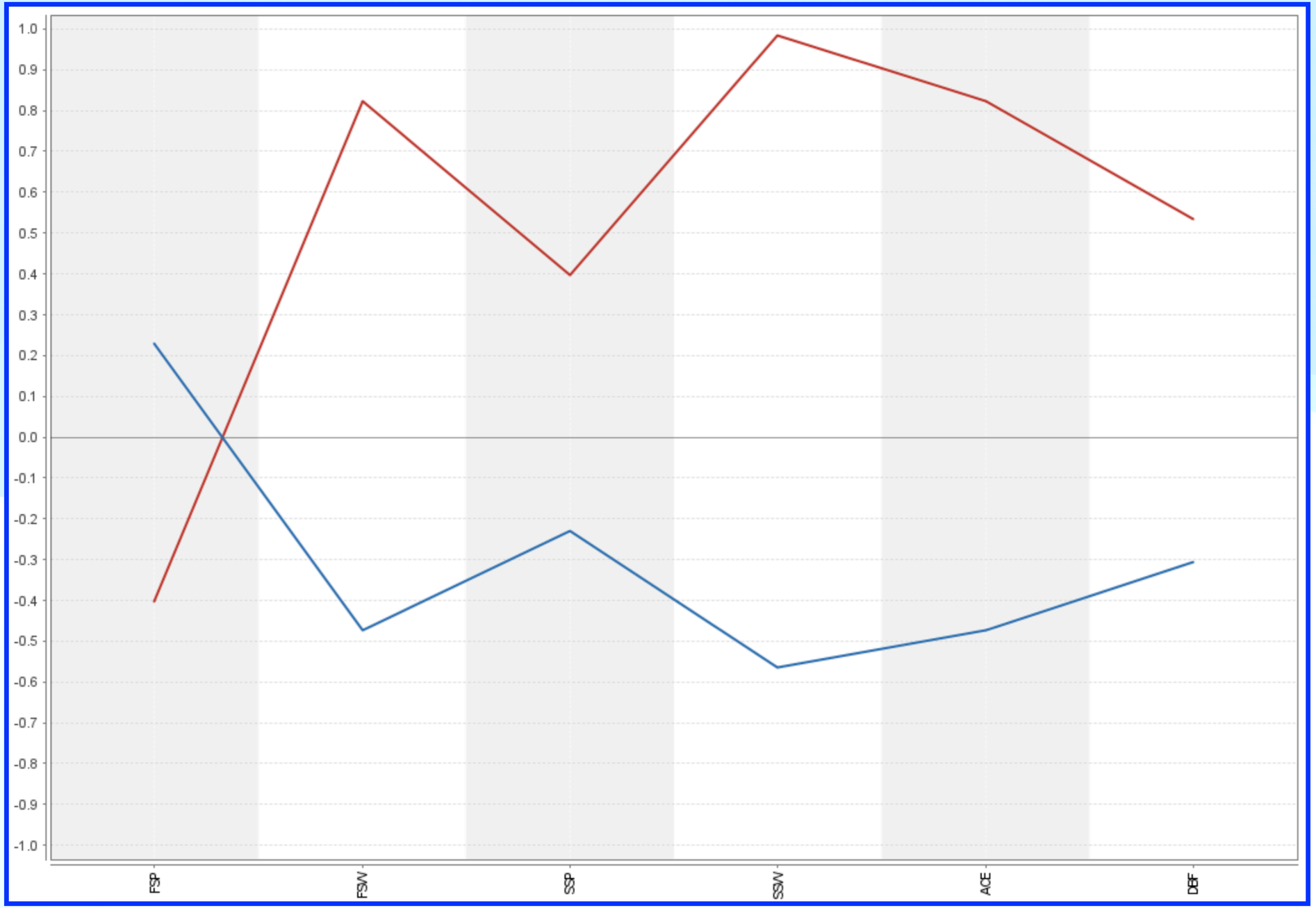
Prediction results

- Neural network: 65.38%
- SVM: 69.23%
- Limiting dataset size

Data analysis in tennis

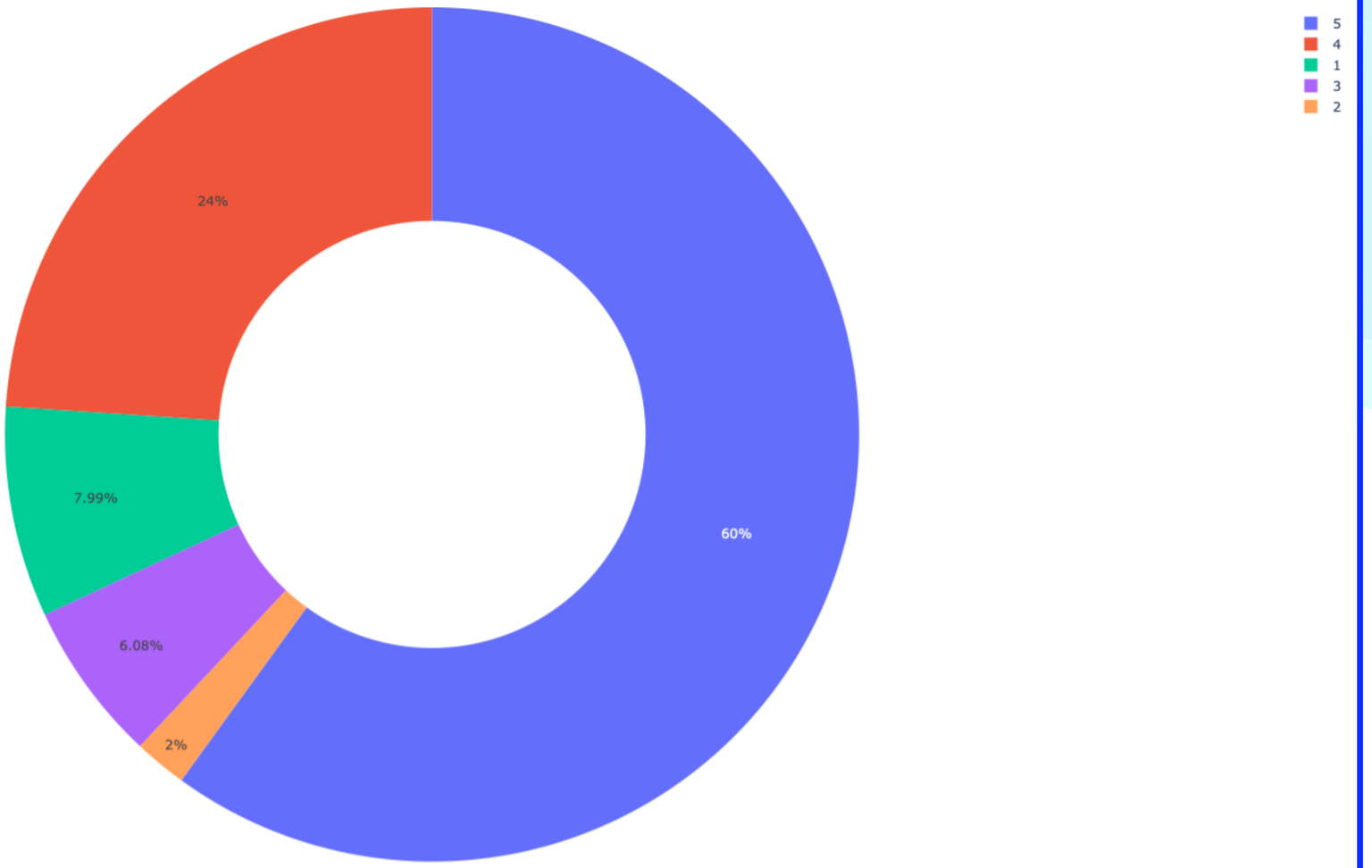
Clustering players by serve

- Data relating to submissions only
- Clustering method k-means
- 2 clusters


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
FSP
FSW
SSP
SSW
ACE
DBF

Sentiment analysis of product reviews

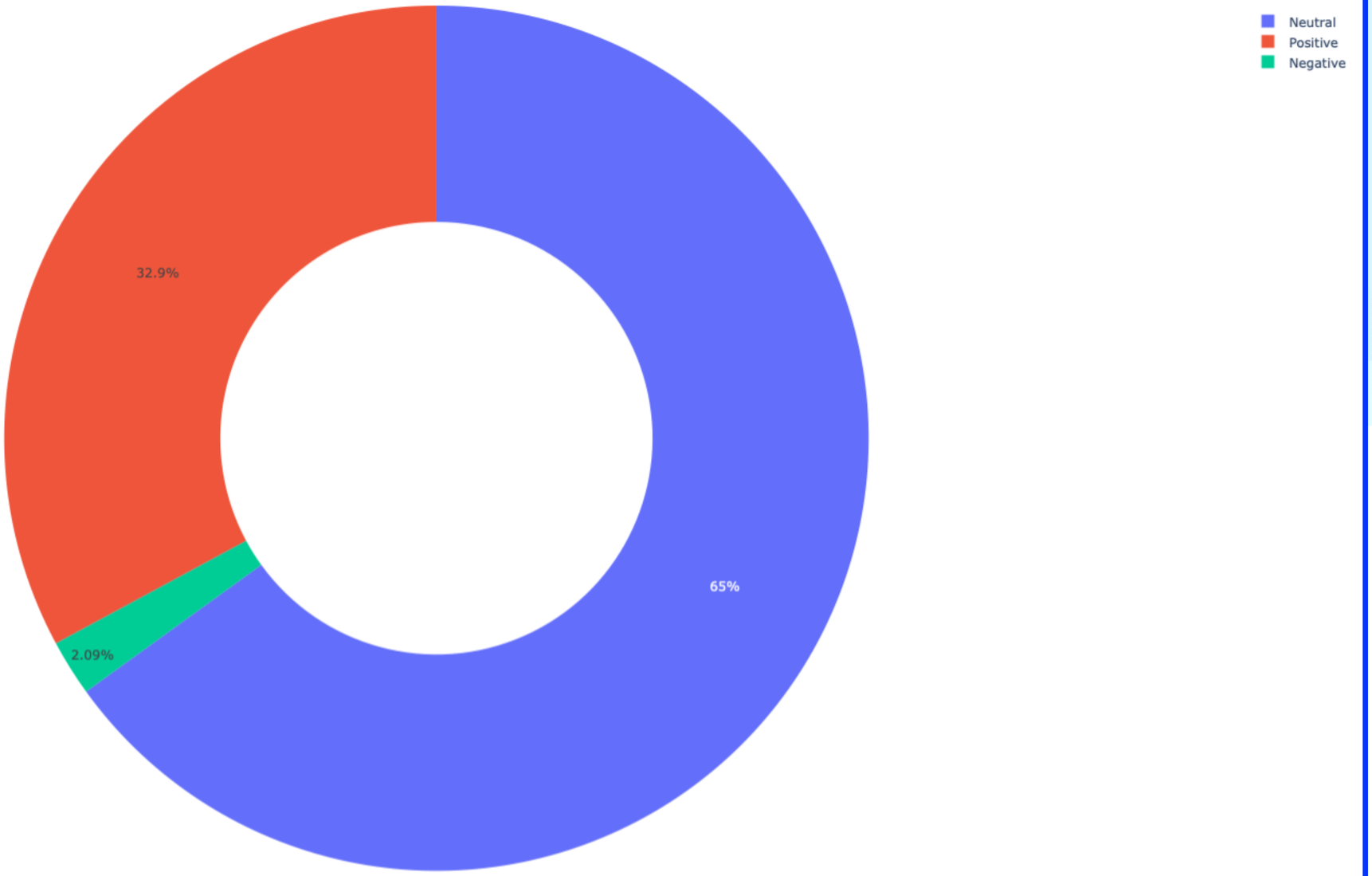
- Dataset (product, reviews and ratings)
- Review cleanup (stop words, stemming)
- Visualizing reviews using word cloud


5
4
1
3
2
24%
7.99%
60%
6.08%
2%


good
phone
use
nice
look
best
camera
display
great
perform
tv
awesom
mobil
product
valu money
work
one
monitor
price rang
samsung
better
buy
price
go
love
time
flipkart
day
cool
perfect
super
need
feel
realm
devic
refriger
laptop
smooth
wash
featur
charg
give
game
thing
superb
thank
realli
qualiti good
camera qualiti
screen
excel
come
compar
machin
sound
speaker
connect
display good
recommend
hour
make
power
fast
much
issu
well
instal
purchas
camera good
decent
first
brand
model
got
low
expect
ok
bought
amaz
happi
set
even
good product
mi
overal
bad
hour
front camera
good batteri
product good
thank flipkart
pictur qualiti
batteri life
batteri backup
design
wash machin
batteri
nice product
deliveri
bright
sound qualiti

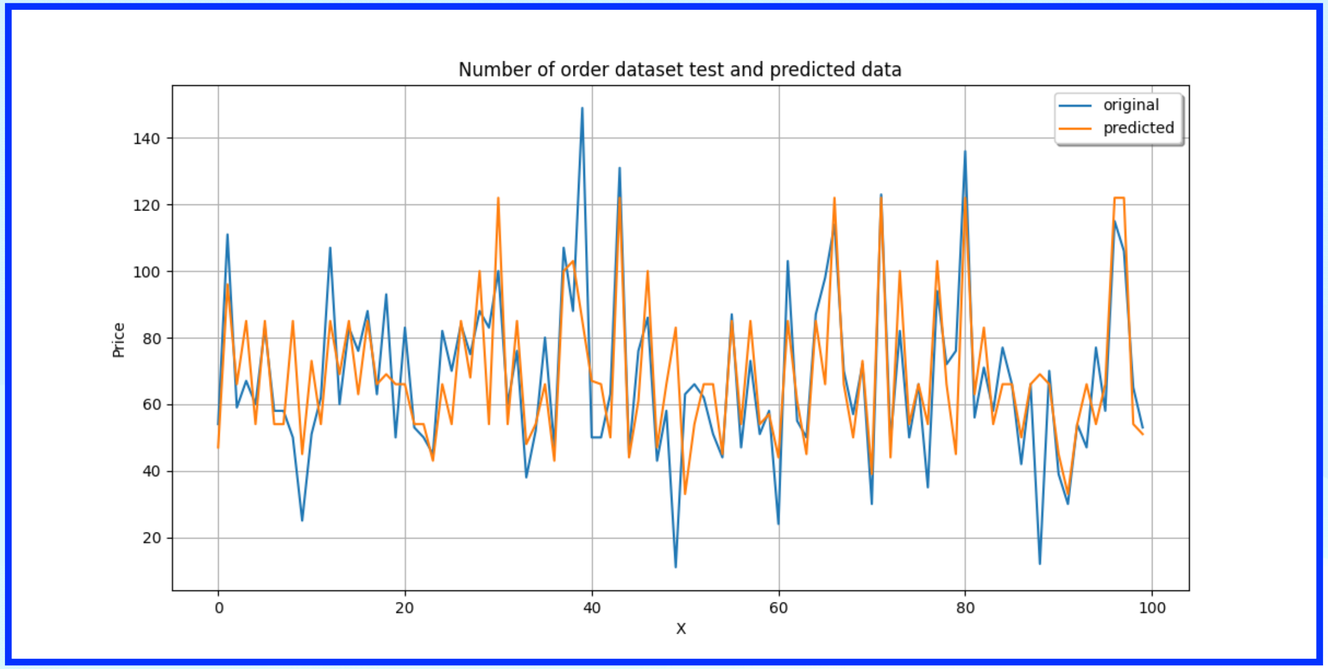
Sentiment analysis of product reviews

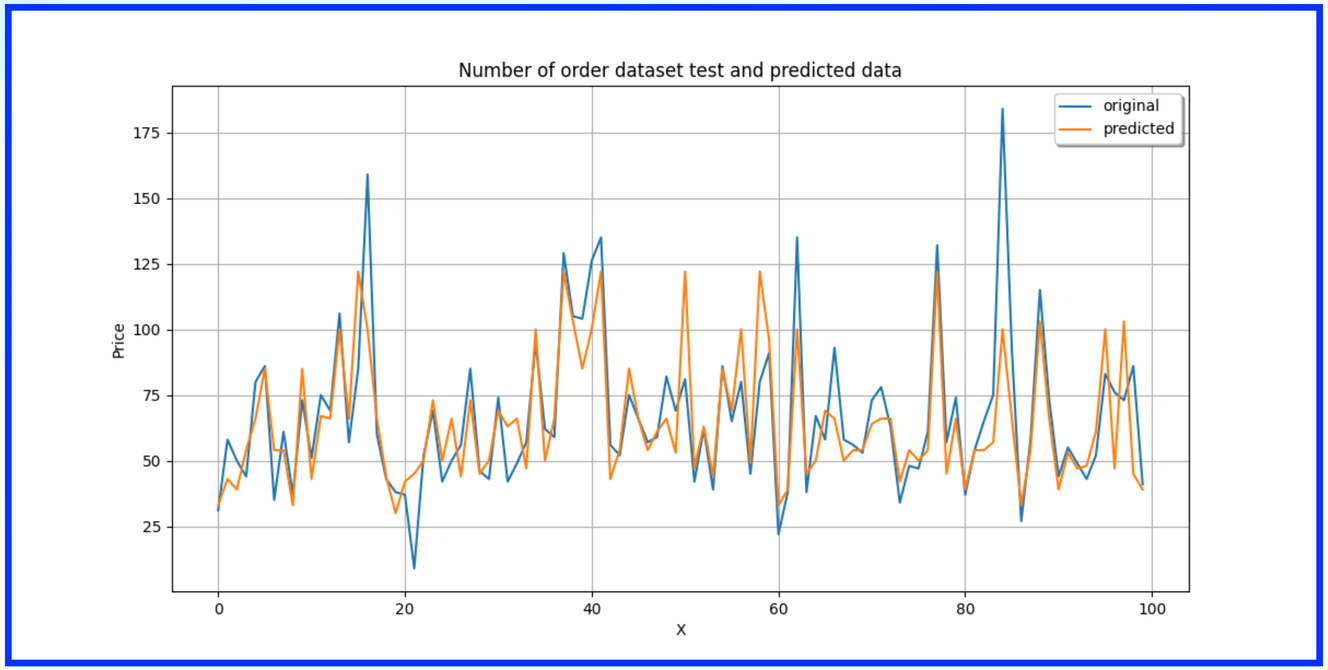
- Natural Language Toolkit
- Vader lexicon
- positive, neutral, negative and composite scores of each review


Neutral
Positive
Negative
32.9%
2.09%
65%

Prediction of the number of orders

- Dataset (Store_Type, Location_Type, Holiday, Discount and number of orders)
- Numeric values only
- To train the model, we use the light gradient boosting regression algorithm (LGBMR)
- On average, the number of orders varies by 13.33
- Root mean square deviation: 19.51


Number of order dataset test and predicted data
original
predicted
Price
X
0
20
40
60
80
100
20
40
60
80
100
120
140


Number of order dataset test and predicted data
original
predicted
Price
X
0
20
40
60
80
100
25
50
75
100
125
150
175

Click-through rate prediction

- Dataset (time on page, age, income group, daily internet usage, gender, click-through)
- Logistic regression
- Accuracy: 71.37%
- F1 score: 0.7074