

A RESEARCH STUDY OF SELECTING APPROPRIATE MACHINE LEARNING ALGORITHMS TO DEAL WITH THREE POPULAR MACHINE LEARNING PROJECTS

DATA MINING & MACHINE LEARNING 1 CA

STUDENT NAME: JIA LIN

STUDENT NUMBER: 22117644



DATASETS AND ML ALGORITHMS



HOUSING PRICE PREDICTION

LINEAR REGRESSION

DECISION TREE REGRESSOR

RANDOM FOREST REGRESSOR



BANK PRODUCT SALES

LOGISTICAL REGRESSION

DECISION TREE CLASSIFIER

KNN CLASSIFIER



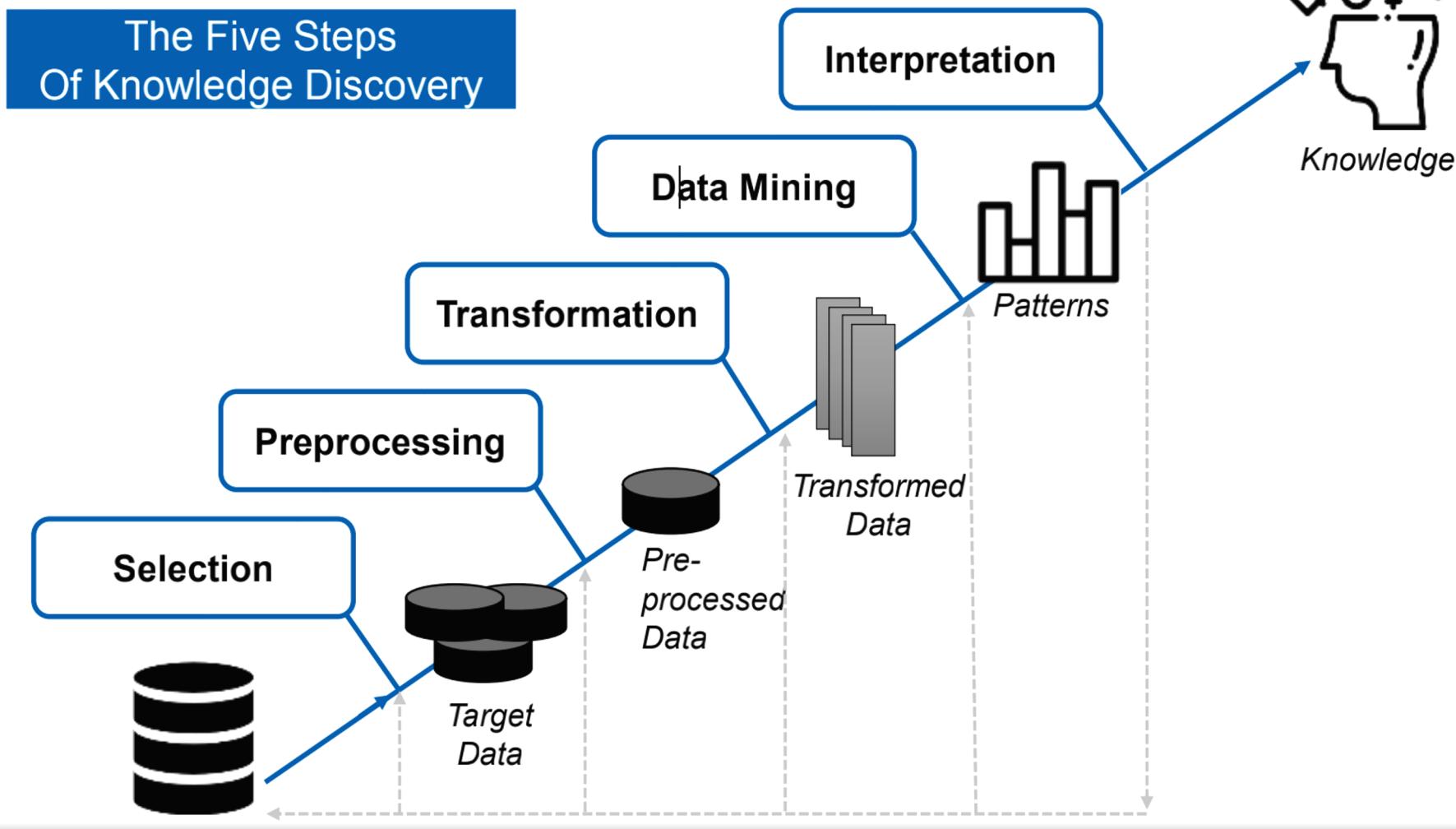
SENTIMENT ANALYSIS

NAÏVE BAYES

RAPIDAPI.COM

KAGGLE.COM

The Five Steps Of Knowledge Discovery



DEMO

jupyter JiaLin_x22117644_DMML1_CA_Dataset1_Hous... Last Checkpoint: Last Monday at 15:27 (autosaved)  Logout

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipykernel)

File Cell Kernel Help Trusted Python 3 (ipykernel)

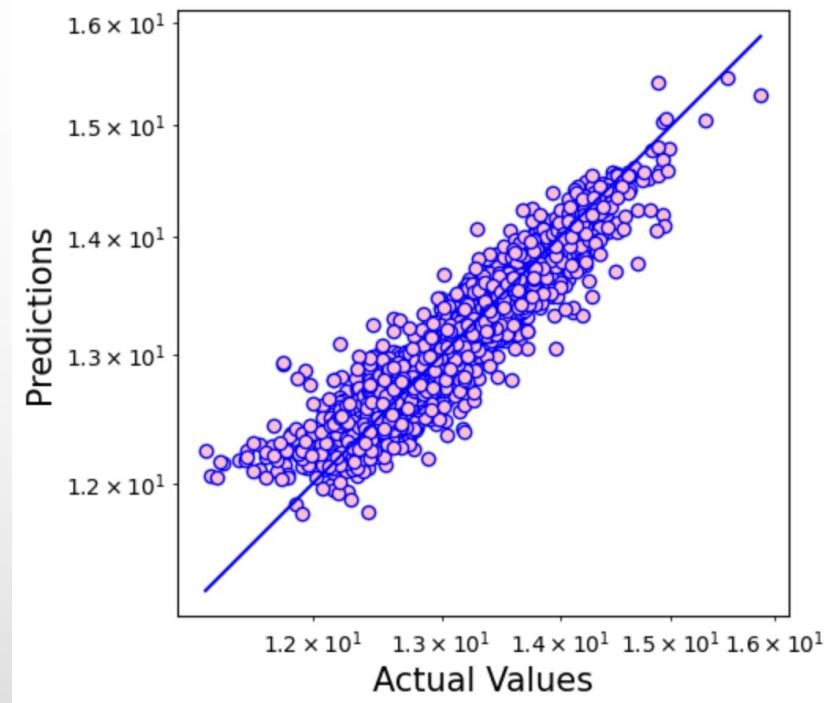
1 Dataset1 Housing Price Prediction

1.1 Import Required Packages

```
In [1]: 1 import pandas as pd  
2 import numpy as np  
3  
4 import seaborn as sns # boxplot, histplot  
5 import matplotlib.pyplot as plt # plot  
6  
7 from sklearn.feature_selection import SelectPercentile # Feature Selection  
8 from sklearn.feature_selection import f_regression # for linear regression  
9  
10 from sklearn import preprocessing # Scaling IV Normalisation  
11 from sklearn.preprocessing import StandardScaler # Scaling IV Standardising  
12  
13  
14 from sklearn import model_selection
```

EVALUATION OF REGRESSION MODELS

	linear	DecisionTree	RandomForest
Accuracy	0.765766	0.776085	0.882892
MSE	0.066952	0.064002	0.033473
RMSE	0.258750	0.252987	0.182957
MAE	0.200053	0.181223	0.128985
R ²	0.765766	0.776085	0.882892
Adjusted R ²	0.764831	0.775190	0.882425



DMML1/CA/

JiaLin_x22117644_DMML1_CA_...

JiaLin_x22117644_DMML1_CA_D...

DEMO

jupyter JiaLin_x22117644_DMML1_CA_Dataset2_Bank... Last Checkpoint: 16 hours ago

File Edit View Insert Cell Kernel Navigate Widgets Help

Trusted

Python 3 (ipykernel)



1 Dataset2 Bank Product Sales Forecast

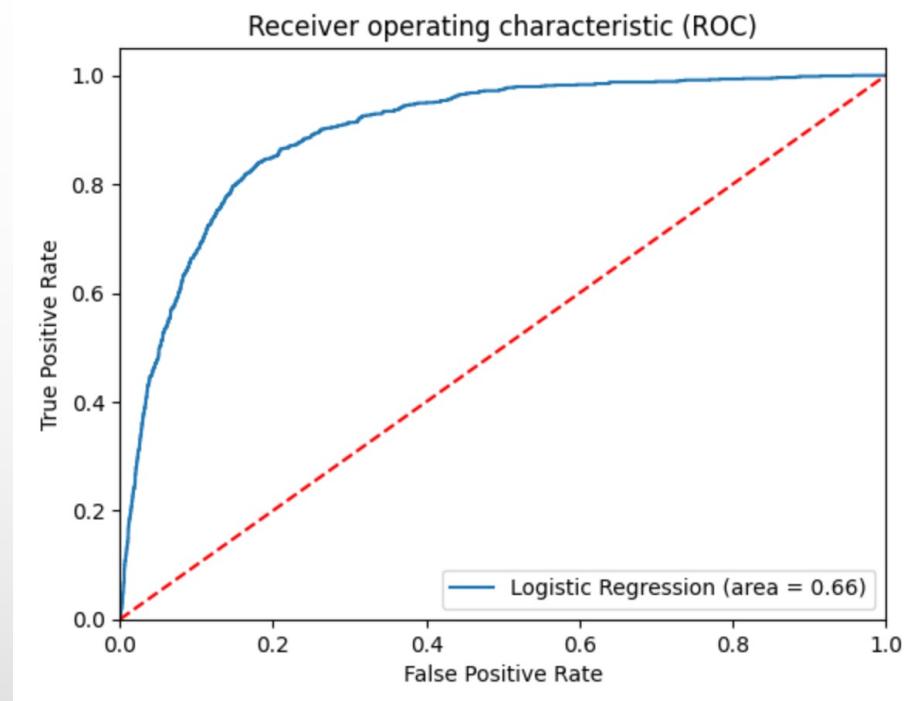
1.1 Import Required Packages

In []:

```
1 import pandas as pd
2 import numpy as np
3
4 import seaborn as sns # boxplot, histplot
5 import matplotlib.pyplot as plt # plot
6
7 from sklearn.feature_selection import SelectPercentile # Feature Selection
8 from sklearn.feature_selection import f_regression # for linear regression
9
10 from sklearn import preprocessing # Scaling IV Normalisation
11 from sklearn.preprocessing import StandardScaler # Scaling IV Standardising
12
13 from sklearn import decomposition # Principal Component Analysis
14 from sklearn import tree
```

EVALUATION OF LOGISTICAL REGRESSION

	precision	recall	f1-score	support
0	0.92	0.97	0.94	7980
1	0.62	0.35	0.45	1063
accuracy			0.90	9043
macro avg	0.77	0.66	0.70	9043
weighted avg	0.88	0.90	0.89	9043



DMML1/CA/ JiaLin_x22117644_DMML1_CA_Dataset3_Sentim... JiaLin.

jupyter JiaLin_x22117644_DMML1_CA_Dataset3_Sentim... Last Checkpoint: Yesterday

DEMO

File Edit View Insert Cell Kernel Navigate Widgets Help Not Trusted Python 3 (ipykernel)

File Run Cell Markdown

1 Dataset3 Sentiment Analysis YouTube Comments (Step1)

1.1 Import Required Packages

In [1]:

```
1 import requests          # library request is used to fetch data from API
2 import json               # library json is used load json objects
3 import datetime           # library datetime is used to convert between datetime and
4 from datetime import date
5
6 import pandas as pd
7 import numpy as np
8
9 #!pip3 install textblob
10 from textblob import TextBlob # Natural Language Toolkit (NLTK)
```

executed in 1.06s, finished 02:10:01 2023-05-02

Jupyter JiaLin_x22117644_DMML1_CA_Dataset3_Senti... Last Checkpoint: Last Tuesday at 11:11:16 AM

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipykernel)

DEMO

1 Dataset3 Sentiment Analysis YouTube Comments (Step2)

1.1 Import Required Packages

```
In [1]: 1 from math import log  
2 from bs4 import BeautifulSoup  
3 import string  
4 import re  
5  
6 import nltk  
7 from nltk.stem import WordNetLemmatizer  
8 from nltk.stem.snowball import SnowballStemmer  
9 from nltk.stem.porter import PorterStemmer  
10 from nltk.corpus import stopwords  
11  
12 #nltk.download('all')
```

executed in 665ms, finished 18:11:16 2023-05-05

EVALUATION OF NAÏVE BAYES

```
File Type: <class 'str'>
Finished with neg train 419545
File Type: <class 'str'>
Finished with pos train 1309125
contentsNeg 913481
contentsPos 175136
Vocab 35949
Starting test
Num test comment = 208
The accuracy for negative testing comments is: 92.31 %
Num test comment = 158
The accuracy for negative testing comments is: 92.41 %
Neg test finished
Num test comment = 882
The accuracy for positive testing comments is: 71.88 %
Num test comment = 658
The accuracy for positive testing comments is: 71.58 %
Pos test finished
```

THANK YOU

JIA LIN

X22117644@STUDENT.NCIRL.IE