

Data Science Internship at Lets Grow More

Devrat Yashraj Deepak

yashraj.devrat.aids.2020@vpkbiet.org

Supervisor

Prof. Digambar M. Padulkar



Department of Artificial Intelligence & Data Science
Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology
Vidyanagari, Baramati-413133

- Introduction
- Internship Objectives
- Motivation
- Technology
- Analysis
- Observation/Results
- Conclusion
- Organization Information
- References



Task - 1 (Iris Flower Classification)

- Classification , Logistic Regression

Task - 2 (Stock Market Prediction using LSTM Network)

- Plotting Graphs using matplotlib

Task - 3 (Exploratory Data Analysis on Terrorism Dataset)

- Data Visualization



Internship Objectives

Gaining practical experience

- The internship provides an opportunity to gain practical experience in the field of data science by working on real-world projects and data sets.

Applying theoretical knowledge

- Interns can apply the theoretical knowledge gained through academic coursework to realworld

Developing new skills

- Interns can learn new skills and techniques that are not typically taught in academic coursework, such as how to work with large data sets or how to use specific tools or software.



Motivation

- Internships are helpful for gaining hands-on experience in the domain which we like so I got good opportunity to showcase my skills.
- Eagerness for learning data science.
- Develop interest in the domain.
- To discover the patterns in data.
- To process the collecting, storing, segregating and analyzing data.



Logistic Regression

Type of supervised learning . Probability is in binary event occurring. In Iris flower Classification I used this to classify categorical variables.

Confusion Matrix

I plotted confusion matrix of Iris Dataset . Here Came to about true labels and mislabels . It predicted if the predictions are correct or incorrect.

LSTM Network

I trained the model by taking some input and current cell input , to predict future stock market prices . I used this network to find hidden patterns and generate high accuracy by running epochs . This network is good for complex datasets .



Strategy

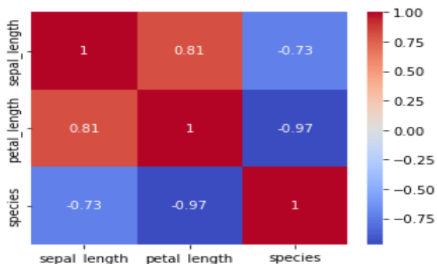
- 1 Data Preparation
- 2 Feature Selection.
- 3 Model Selection.
- 4 Model Training and Evaluation



Iris Flower Classification

```
In [33]: corr = data.corr()  
fig, ax = plt.subplots(figsize=(5,4))  
sns.heatmap(corr, annot=True, ax=ax, cmap = 'coolwarm')
```

Out[33]: <AxesSubplot:>



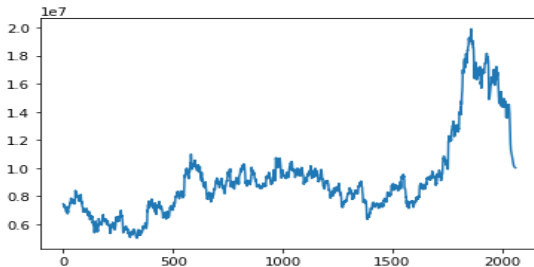
Heatmap Plot



Stock Market Analysis using LSTM Network

```
In [ ]: plt.plot(df3)
```

```
Out[85]: [<matplotlib.lines.Line2D at 0x7f764ffa9f10>]
```

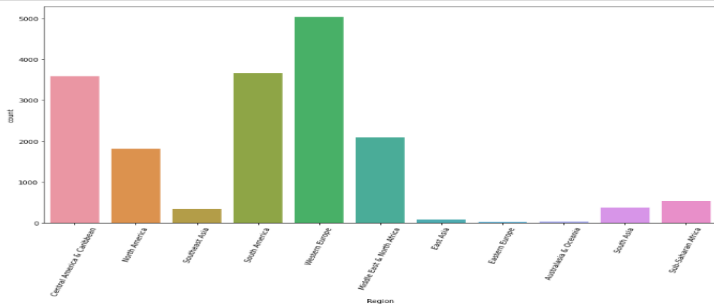


Analysis



Exploratory Data Analysis on Terrorism Dataset

```
In [13]: #REGION AFFECTED BY TERRORIST ATTACK
dataset['Region'].value_counts()
plt.figure(figsize=(15,8))
sns.countplot(x='Region',data=dataset)
plt.xticks(rotation=70)
plt.show()
```



Analysis



Observations/Results

1. Accuracy :

Model Used : Logistic Regression

Task 1 - The accuracy of the model is 97.37 percent.

2. Accuracy: -

Model Used : LSTM Network

Task 2 - The accuracy of the model is 90 percent.



Advantages

- This internship will surely help me to build my portfolio and increase chance of getting selected in MNC.
- Python for Data Science is a powerful tool for analysing and visualising complex datasets.
- More features than regular Microsoft Excel.
- Helpful in Machine Learning to train and test the data .
- Recognition of an intern of company due to certification .



Disadvantages

- No physical experience of internship.
- It requires lot of knowledge to handle big data.
- Statistics is more important in data science so one needs have good knowledge of it.
- Need to train the models so it is time consuming.
- Essential knowledge of Python or R language.



This internship consisted of various projects such as iris flower classification, stock market prediction , and exploratory data analysis on Terrorism dataset. The analysis of these projects highlighted the effectiveness of machine learning algorithms in solving complex problems and generating insights. The internship helped in developing skills such as data wrangling, data preprocessing, feature engineering, model selection, and evaluation.



- [1] <https://data-flair.training/blogs/iris-flower-classification/>
- [2] [https://www.analyticsvidhya.com/blog/2020/08/lstm network
-using-machine-learning-dataset](https://www.analyticsvidhya.com/blog/2020/08/lstm-network-using-machine-learning-dataset)
- [3] [https://thecleverprogrammer.com/2021/07/12/stockmarket - analysis](https://thecleverprogrammer.com/2021/07/12/stockmarket-analysis)
- [4] Python for Data Analysis-Book



- Platform : Operating System: Windows 7 or above, Ubuntu 12 or above
- IDE: Jupyter Notebook, Google Colab Notebook
- Programming Language : Python





- Company is located at Allahbad Which work in majorly in Data Science Domain . It always open to connecting new people among them over this journey of endless learning.
- Services offerd web development , Android Development , Machine Learning , Data Science etc. One of the leading company in Data Science field.



Thank You

