Final Project Work 4Geeks Academy 2022

Group 1: Salvatore, Carlos, Concalo

Time Series and NLP for Market and Economic Predictions

Time Series and NLP for Market and Economic Predictions

Part 1

Nowadays, **Time Series Prediction** or **Time Series Forecasting** represents a critical task for the Companies, since it has allowed some businesses to survive such as Zalando. Zalando has 4 AI Labs in Europe at moment:
https://engineering.zalando.com/posts/2016/10/deep-learning-for-understanding-consumer-histories.html.
Also Amazon optimizes its production thanks to Machine Learning and Deep Learning Models. For many of these companies it is getting important the equation **model = money**, since a lower prediction error can represent an higher earning margin. Therefore, finding the best model for a certain task is remarkable and can mean non-loss money or gain. In this Final Project we propose to build a model to predict **Stock Exchange by exploiting news**, taking some Kaggle competition as example: https://www.kaggle.com/c/two-sigma-financial-news. News have a huge predictive power and can be harnessed to forecast financial results and generate economic impact. On a side, it is necessary to create one or more models for Financial Time Series Prediction on the other side it is required to build the best models in terms of error (MAE, MAPE, ecc). Furthermore, we require the creation of Deep Learning models for the text classification of news such as **Thomson Reuters** which must be filtered on one stock (i.e. apple, etc) and encoded with Word Embeddings. Is is possible to have a look at this on how to download and arrange data: https://www.kaggle.com/code/danofer/two-sigma-financial-news-export/notebook.
Try to investigate the accuracy of traditional Deep Learning Models (i.e. LSTM, BiLSTM, Recurrent ConvNets) or more cutting edge models (i.e. Transformers, Bert, etc.).
Adopt MAPE metric for the Time-Series Prediction, Accuracy and F1-Score for the classification problem and a Profitability Measure (design it or find one algorithm on the web) to examine the brekthroughs in terms of revenues of the model.
Make **EDA** (Exploatory Data Analysis) and plot output panels plotting: training curves, classification barplots, misclassifications, cluster analysis and all the due analysis.

Part 2
Adopt the best ML/DL classification models in order to make **Loan Approval Prediction** and **Fraud Detection**. For the Load Prediction start from this:
https://www.analyticsvidhya.com/blog/2022/02/loan-approval-prediction-machine-learning/
and/or this:
https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset
For the fraud detection take in consideration this Transactions Fraud Detection:
https://www.kaggle.com/c/ieee-fraud-detection
Make **EDA** (Exploatory Data Analysis) and plot output panels plotting: training curves, classification barplots, misclassifications, cluster analysis and all the due analysis. The purpose of this Final Project Work is to show the way Artificial Intelligence can be useful in Economy.