

Solution Sheet

1. Which model have you used for stock price prediction? Explain your model.

We have worked on a dataset called “train_dataset” which was about Stock Price prediction. The dataset which we used had a target column which means it was a supervised learning problem. So, we have built a **regression model** for this problem.

The **challenges we faced** while working on this project were:

- The dataset had a lot of missing values. We tried filling the missing values with mean, median, KNN Imputer but the best accuracy we got was with KNN Imputer. So, we chose to fill missing values with KNN Imputer.
- The dataset **didn't have a lot of outliers** but there were few which we handled using Inter quartile range (IQR).
- We used **Pearson Correlation Technique** and we plotted the **Heat Map** to find the correlation between the features. It was found that Index and General index

column were highly correlated with each other. So, we removed General index feature.

- We plotted **Scatter Matrix** to find that how features were related to the target feature and we found that two features to be skewed with respect to the Stock Price. So, we used the $\log(\text{Stock Price})$ feature.
- There were **few categorical columns**. To handle those, we used **Dummy Encoding**.
- We also did **feature scaling** because (normalized the features) because as we were building the Linear Regression model and Linear Regression model uses gradient descent to find the parameters and the gradient descent would converge faster to global minimum if the features would be scaled.

CONCLUSION: (RESULT)

After this, we built our regression model and got a pretty good accuracy of **88.5% (approx. 89%)** which means that the model we used was good.

We evaluated the mean squared error of the model which came out to be 0.09.

At the end, we took the exponential value of the Predicted Stock Prices to convert the log value of Stock Price to actual Stock Price values.

2. Which model have you used for Put-Call ratio Time series prediction? Explain your model.

We have used **Recurrent Neural Network Model** to do Time Series Prediction because the output from previous step are fed as input to the current step. The Put-Call ratio of a stock is a time-dependent parameter, for which we had to come up with a **Time-series prediction model**.

We have used LSTM model to do time series prediction as well.

It is special kind of recurrent neural network that is capable of learning long term dependencies in data. This is achieved because the recurring module of the model has a combination of four layers interacting with each other.