**Meeting-20231211\_213540-Meeting Recording**

December 11, 2023, 2:35AM

15m 14s

 **Vemula,Vijval** 0:04  
What's that it.

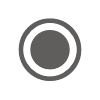
 **Patil,Priyanka** 0:05  
Hello all we are presenting Apple stock prediction.  
We are member of three which will Tyler and myself Priyanka.  
To uh, these are the content of our slide in the upcoming slide.  
You can see the not knowing the dataset.  
Exploratory data analysis then proposed machine learning models.  
Then they're comparative analysis of the algorithmic results, conclusion, limitation and future work you.  
Still in the data set we have 10,840 rows and seven columns.  
There was no.  
There were no missing values in the data set and like we have analyzed the stock prices from 1980 to till date and the source of our data set was Yahoo Finance.  
Uh API and this is the static.  
OK.  
So in India part we have analyzed historical Apple stock price from 1980 till date and we have gained insights into long term trends volatility and it's patterns.  
Then we have identified events that impacted Apple stock prices and like we have a we tried to able to understand the relationship between Apple stock and broader market.  
Then we visualize Apple stock prices over the time using charts.  
Uh, so we have detected seasonal patterns or recurring trends and then we derive key insights to inform trading, modeling and prediction.  
So this is the graphical representation of the stock prices over the time.  
So we as we can see a till date the highest was 200 USD.  
So for a the the.  
So we have proposed A3 machine learning models based on our research and those are linear regression, random forest classifier and long short term memory algorithm.  
So in linear regression, we split the data into test and training sets and then train the model with the training set to predict the stocks future price.  
Then for random forest classifier what we have done is we have created a binary target variable that tells whether tomorrow's stock price will be greater than or less than two days closing price.  
And this is the hypothesis of our second model.  
And for last LSTM, we built a model that has neural network layers.  
So the main aim of this model is to predict tomorrow's stock price using the past 60 days of data of stock price.  
So now next I will hand over to Tyler.

 **Vemula,Vijval** 3:36  
Uh, I'll be taking it.  
Sorry.

 **Patil,Priyanka** 3:38  
OK.

 **Vemula,Vijval** 3:39  
Uh, so this is our first model linear regression algorithm.  
So we basically wanted to know how regression and classifier works for time series analysis and the first thing that popped in our mind is linear regression because it is simple and easy to understand.  
So uh, we just thought that the stock price relation would be a linear between time and the stock price, and in the context of stock prediction, simplicity can be an advantage, especially like when dealing with small to moderate amount of data.  
And these are the a metrics for our inner linear regression model.  
UM, by looking at these metrics, we can just tell that uh, the linear regression model is not that much up to the mark and this is the graphical representation.  
The graph is completely skewed.  
The predicted values and actual values are not at all matching to each other, and even the residuals of those actual prices and predicted prices are not even closer to 0.  
If you dual line, so we can say that out of all three this has the least performance.  
Linear regression has the least performance coming to the next model.  
This is the random random forest classifier.  
The reason why we are telling this is a classifier is because we have created a binary variable called target and that target tells whether the stock price of tomorrow's uh stock is greater than or lesser than the closing price of today's stock price.  
So if tomorrow's stock price is lesser, it'll be 0.  
If it is greater, it'll be one and our aim is to predict the target column.  
And this is the overall report of our uh uh model.  
We have an accuracy of 51% which is slightly better than linear regression.  
We are not selling that this is completely better, but this is slightly better than linear regression and there clearly appears to be an imbalance in the data set as indicated by the difference in precision, recall and different score between the two classes.  
And this is the uh graph for predictions and target values.  
So we can just tell that if these two lines overlap completely, or at least closely, we can say that our model is very good.  
But unfortunately only 51% of the uh probability has been matched here.  
So this is a better model compared to linear regression and the third one is the long short term memory algorithm.  
It completely works on neural networks and this is very simple.  
The name itself fares long short term memory, so we'll just take the stock prices or the closing values of previous sixty days stocks and then we'll predict the price of stock on the 61st day.  
So the architecture of the model is very simple.  
We have two layers of dents.  
Uh neural networks and two layers of LSTM neural networks.  
The first two are LSTM.  
The last two layers are dense, so we have 4.  
Uh, a neural layers and in the first layer we have 100 neurons.  
In the second layer, we have 75 neurons in the third layer we have 15 neurons and in the fourth layer we have only one neuron, which will contribute to the output and things neural network needs uh data.  
That is not very much biased, so we have truncated the data from 2012 to till date like we have deleted from 1980 to 2011 the the data related in that time period because it is very much flat during that time period and one main disadvantage of this model is the values should be in the range of 0 to one.  
So we we kind of a scaled the train values Training fit values using min Max scaler 2 zero to 1 and then we predicted the predictions and then we in applied an inverse transform to those prediction values so as to bring those predictions to the actual scale.  
And this is the result for the predictions.  
As you can see, it is very much closer to the actual price, which is why test and this is the RMSE error.  
It is very much low, but we thought like we need to make some more changes like we need to bring these two lines closer.  
So what we have done is we have added the absolute value of the mean of the difference between these two lines like predictions and white test.  
We added that to the predictions and we got a more clear and completely, I mean not completely 80% overlapping.  
Graph and this furthermore decreased the RMSE to a great extent.  
And this is the graphical representation of our train test and uh prediction values.  
So the blue one is the uh Training values and whereas the orange one, the actual test values and green is the predictions and you can see orange and green are overlapping almost with each other, which means that LSTM is our hero right now.  
Uh, so bit did alleged Ms outperformed previous two models.  
So from here, uh Tyler will be taking over.

 **Brenner,Tyler** 10:47  
So we did a.  
So to sum it up, a comparative analysis of out of the algorithmic results for the linear regression, the strengths that we had with that model where we did have a high R-squared, which was .9 indicating that it was a good fit.  
However, the root mean square error and the mean absolute error values suggested that there was a substantial error when predicting those stock prices, and then the concerns with the linear regression model was that due to the high error metrics within the the root mean square and the mean absolute error, it might indicate that there's some limitations in capturing the underlying patterns within the data within the random forest classifier.  
The strengths within that the accuracy was a .51, which indicated that the performance was slightly better than random chance.  
But the precision for class one was relatively high.  
It was a .68, so the concerns with this was that the model appears to struggle with recall for class one, which indicated a high number of false negatives.  
And then for long short long short term memory, the strengths for that model we had an extremely low root mean square error which was 6.29 \* 10 to the negative 7th, which indicated that the long short term memory model made very accurate predictions, at least on the provided data.  
But the concerns with that that model were that the extremely low root mean square error may be a signal of overfitting.  
So it's crucial to assess the models performance on new unseen data to ensure that there's generalization.  
In conclusion, the linear regression model was not very accurate.  
It needs extensive hyper tuning of the parameters.  
The random forest classifier model was a comparatively good performer when compared to the when compared to the linear regression model and then the the long short term memory model was giving the most precise results.  
If it was trained well.  
The limitations for this were that the financial markets, including the stock market, there are all influenced by a myriad of factors of some of them include economic conditions, global events, political developments and investor sentiment.  
But the inherent volatility and unpredictability of these factors make accurate predictions challenging.  
And again, this is just an initial idea.  
Please don't use this for actual prediction purposes and compliance with financial regulations and ethical considerations is very crucial in financial modeling, so those predictions should be made transparently and ethically, with awareness of potential legal implications.  
Some future work to design a system that could handle real time data updates and predictions, because this is crucial for applications that require up to the minute information for different decision making.  
Also, to build an unseen ensemble models that combine predictions from multiple models, such as a combination of long short term memory networks, random forests, or even gradient boosting, boosting algorithms, ensemble models often actually lead to improved generalization and robustness.  
So that's why that would be a good model.  
Also, we consider more complex deep learning architectures such as attention mechanisms or even transfer transformer based models which have actually shown success in sequential data tasks.  
These architectures may actually be captured, may able to be captured it dependencies in time series data.  
And lastly, we can implement a robust evaluation strategy that actually includes periodic model retraining and assessment of model performance on recent data.  
Because the final financial markets evolve and the model effectiveness may change over time.  
Thank you guys very much for watching this and that's all.

 **Brenner,Tyler** stopped transcription