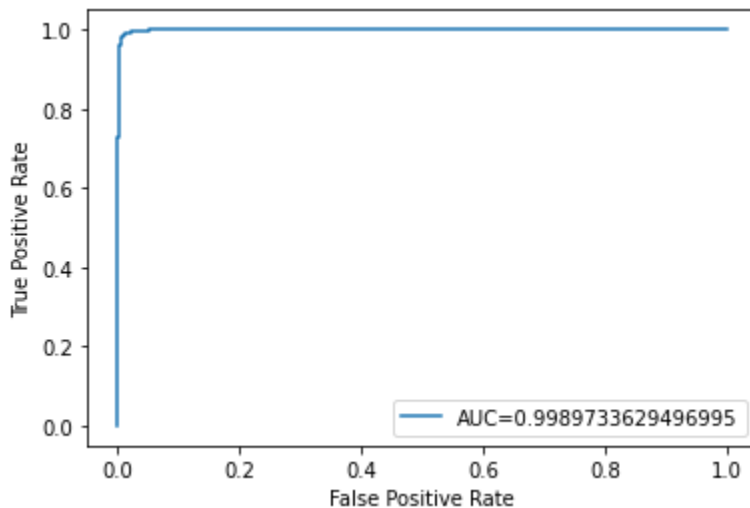


Phase 3 - Report
Arnav Gupta, 200968030, Batch 3
Fake News Detection

- **Model 1 – Logistic Regression**

ROC Curve



According to this visualization the model is overfitting

Logistic Regression Model Performance

Accuracy	98.81%
Precision	98.94%
F1 Score	98.75%

- **Model 2 – Long Term Short Memory Model (LSTM)**

Activation Functions used – ReLu, Sigmoid

Optimizer – Adam

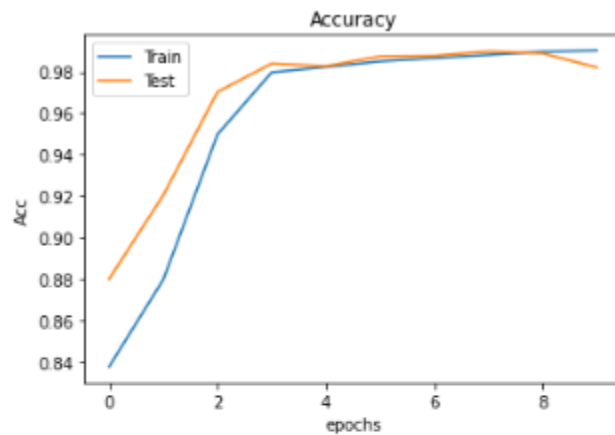
Loss Function – Binary Cross-entropy

Hyperparameters - epochs = 10, batch_size = 199, validation_split = 0.35

The model gives an accuracy of 99%

Test Accuracy of LSTM Model – 98%

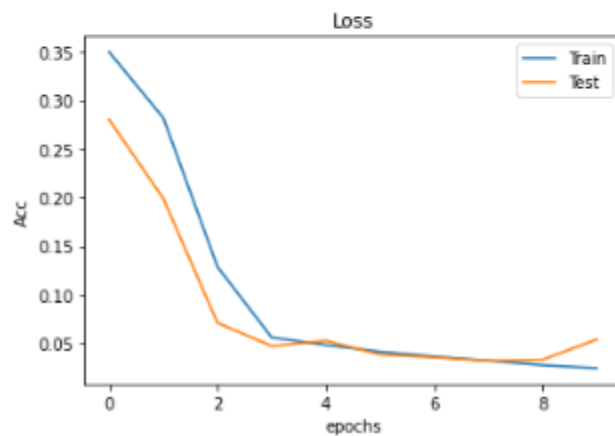
Visualization



Accuracy Curve –

Since test and train are close to each other model is performing well

Loss Curve



- **Model 3 – Recurrent Neural Network Model (RNN)**

Activation Function used – ReLu

Optimizer – Adam

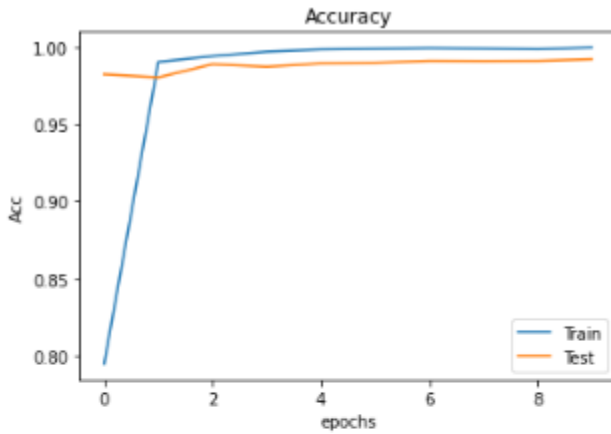
Loss Function – Binary Cross-entropy

Hyperparameters - epochs = 10, batch_size = 69, validation_split = 0.25

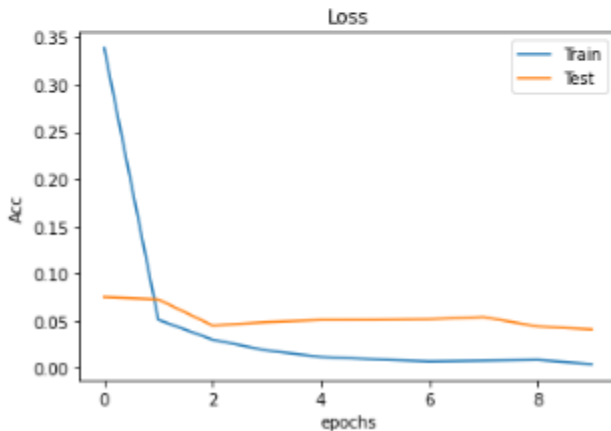
The model gives an accuracy of 99%

Test Accuracy of RNN Model – 99%

Visualization



Accuracy Curve –
Since Train Accuracy > Test Accuracy
the model is overfitting



Loss Curve –
Test Loss > Train Loss

Hyperparameters and their Tuning

The model parameters (eg. weights in Neural Networks, Linear Regression) define how to use input data to get the desired output and are learned at training time. Instead, Hyperparameters determine how our model is structured in the first place.

Optimizer is very important to achieve the possible highest accuracy or minimum loss. There are 7 optimizers to choose from. Each has a different concept behind it. According to this project, Adam optimizer is used.

Adam optimizer involves a combination of two gradient descent methodologies: This algorithm is used to accelerate the gradient descent algorithm by taking into consideration the 'exponentially weighted average' of the gradients. Using averages makes the algorithm converge towards the minima in a faster pace.

Learning rate controls the step size for a model to reach the minimum loss function. A higher learning rate makes the model learn faster, but it may miss the minimum loss function and only reach the surrounding of it. A lower learning rate gives a better chance to find a minimum loss function. As a tradeoff lower learning rate needs higher epochs, or more time and memory capacity resources.

Batch size is the number of training data sub-samples for the input. The smaller batch size makes the learning process faster, but the variance of the validation dataset accuracy is higher. A bigger batch size has a slower learning process, but the validation dataset accuracy has a lower variance.

Validation split helps to improve the model performance by fine-tuning the model after each epoch. The test set informs us about the final accuracy of the model after completing the training phase.

The number of times a whole dataset is passed through the neural network model is called an epoch. Too many epochs will lead to overfitting where the model can predict the data very well, but cannot predict new unseen data well enough. A too-small number of epochs results in underfitting because the neural network has not learned much enough.

Result Analysis

All three models give good accuracy

Logistic Regression gives good accuracy for the fake/real classification task

LSTM - makes it powerful to do forecasting or detecting

Logistic Regression gives an accuracy of 98%

LSTM Model gives an accuracy of 99.02%

RNN gives an accuracy of 99.5%

NLP Fake News Classifier

US/Russian Relations Sour Further As Putin Refuses To Open Snapchat From Obama

0 Add Comment

IN the clearest indication yet that US/Russian relations may be at an all-time low post-Cold War, Russian leader Vladimir Putin is staunchly refusing to open a Snapchat he received from his US counterpart Barack Obama.

"If Putin opens it, it's a poor political play. It shows he's curious, interested in what Obama has to say. By not opening it, he's telling Obama and the world that he has something better to do with his time," explained app and politics expert Morgan Wilde.

A number of Snapchat users have become heavily invested in the ongoing brinkmanship between the two superpowers who are locked in a series of countermeasures in Syria, which have left lives of innocent civilians in the balance and with rumours persisting that Putin hasn't looked at any Western leaders' Snapchat stories in weeks, the

Load random News from test dataset ? [Click here](#).

You can write up to 3172 more words

Predict

FAKE

[Show text processing pipeline](#)

NLP Fake News Classifier

What will Republicans do when they take full control of Congress?

Republicans taking control of Congress this coming week will try to overcome their reputation as a divided party hobbled by infighting by working to reshape policy in ways that Americans will feel in corporate boardrooms, on factory floors and at the gas pump.

Incoming committee chairmen are preparing fresh oversight of federal agencies while rank-and-file members will be encouraged to use a new budget plan and government spending bills to chip away at President Obama's environmental regulations, health-care reform and outreach to Cuba and Iran.

After years of enmeshing with the White House, Republicans are eager to demonstrate productivity and some level of bipartisan cooperation with Obama and the Democrats.

Load random News from test dataset ? [Click here](#).

You can write up to 1873 more words

Predict

REAL

[Show text processing pipeline](#)

Conclusion

LSTM model works best for Fake News Detection due to its memory property. In addition to, overfitting is much less than the other two models of Logistic Regression and RNN. It also gives a significantly good accuracy. LSTM is also good with forecasting.

