Spam Classification in Messages

Kevin Tran, John Meeks, Victor Bialkowski, Ayawavi Touglo, John Ha

Overview

SMS Spam is becoming an increasingly common issue, so finding ways to detect spam messages is extremely beneficial

Our project looks at different ways to detect SMS spam, using a basic MLP, LSTM and BILSTM model

The “SMS Spam Collection Data Set” available from the UCI Machine Learning Repository was used

A model was also ran on the Enron corpus dataset

Dataset Examples

ham What you doing?how are you?
ham Ok lar... Joking wif u oni...
ham dun say so early hor... U c already then say...
ham MY NO. IN LUTON 0125698789 RING ME IF UR AROUND! H*
ham Siva is in hostel aha:-.
ham Cos i was out shopping wif darren jus now n i called him 2 ask wat present he wan lor. Then he started guessing who i was wif n he finally guessed darren lor.
spam FreeMsg: Txt: CALL to No: 86888 & claim your reward of 3 hours talk time to use from your phone now! ubscribe6GBP/ mnth inc 3hrs 16 stop?txtStop
spam Sunshine Quiz! Win a super Sony DVD recorder if you canname the capital of Australia? Text MQUIZ to 82277. B
spam URGENT! Your Mobile No 07808726822 was awarded a L2,000 Bonus Caller Prize on 02/09/03! This is our 2nd attempt to contact YOU! Call 0871-872-9758 BOX95QU

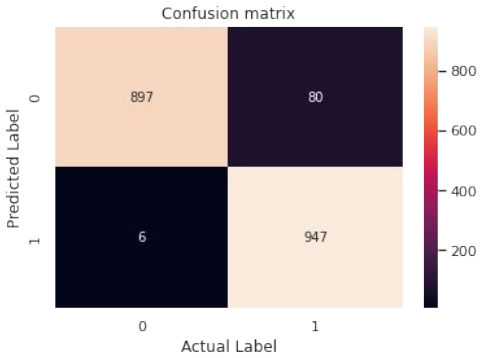
Problem modeling

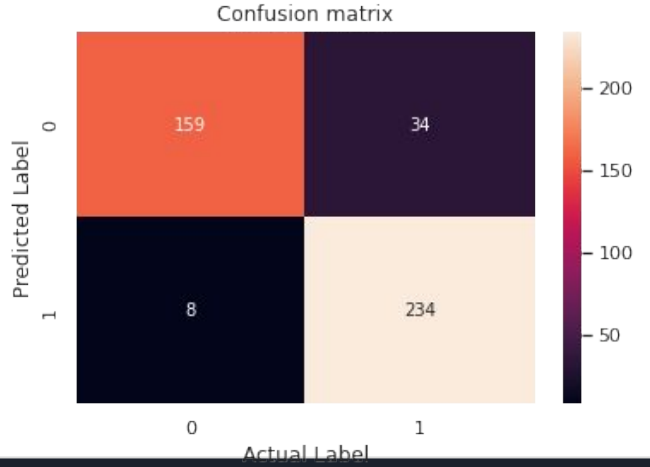
- Type of learning: Supervised Learning
- Input format: A vectorized text array
- Features: Vocab size
- Output format: Value between 0-1 (sigmoid)
- Architectures utilized:
  - MLP
  - BILSTM
  - LSTM

Problem parameters

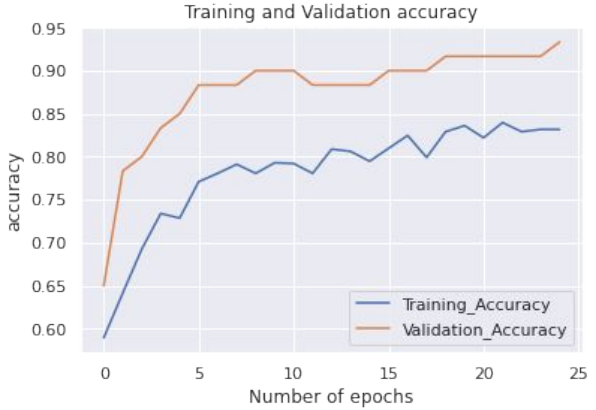
- Loss function: Binary cross entropy
- Duration:
  - Batch size = 32
  - Epochs = 25
- Optimizer: Adam
  - LR = 0.0001

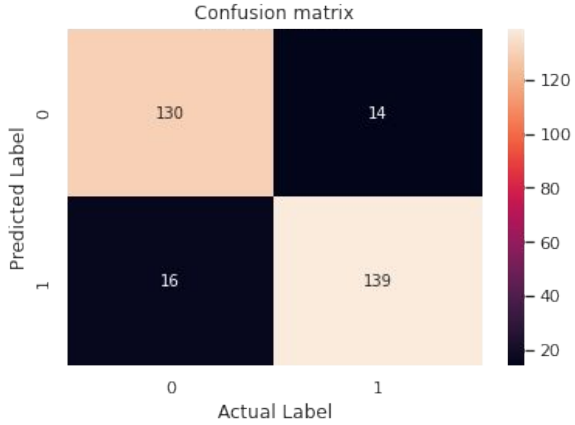
Performance Analysis - MLP Upsampling



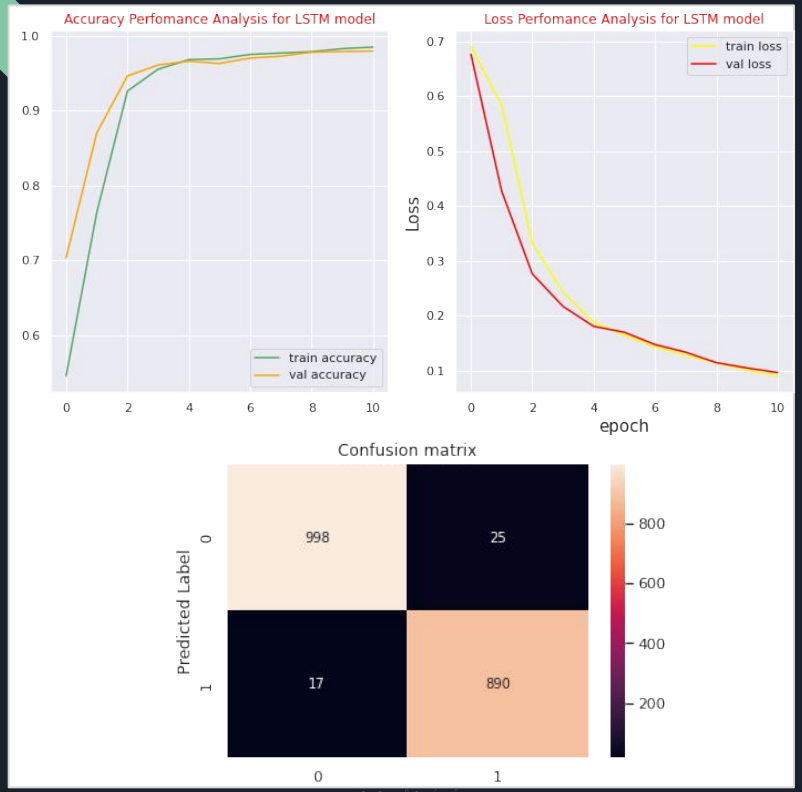


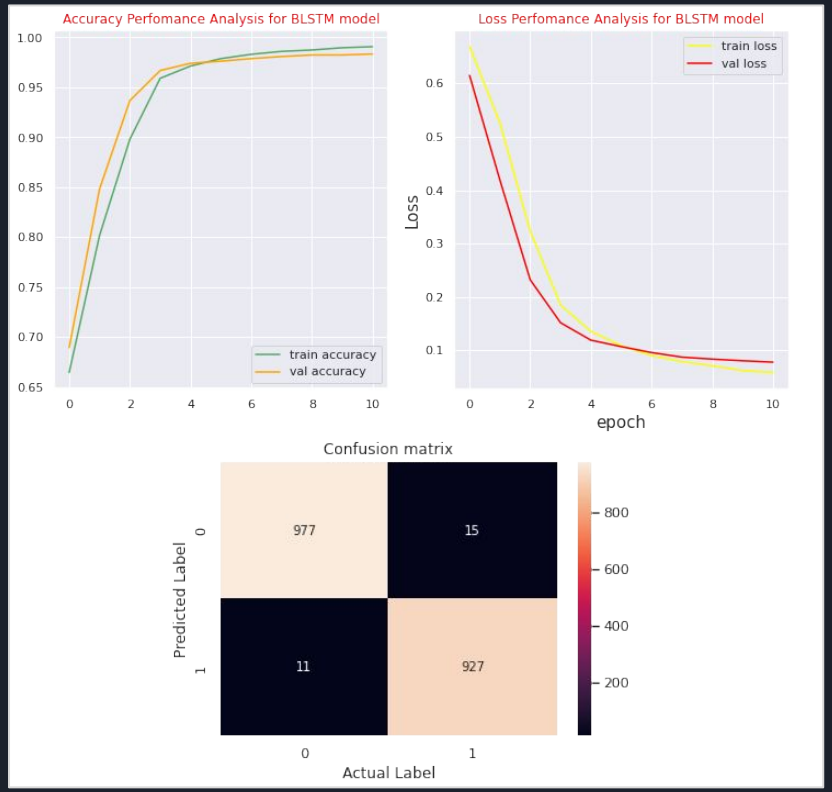
Performance Analysis - MLP (Downsampling)



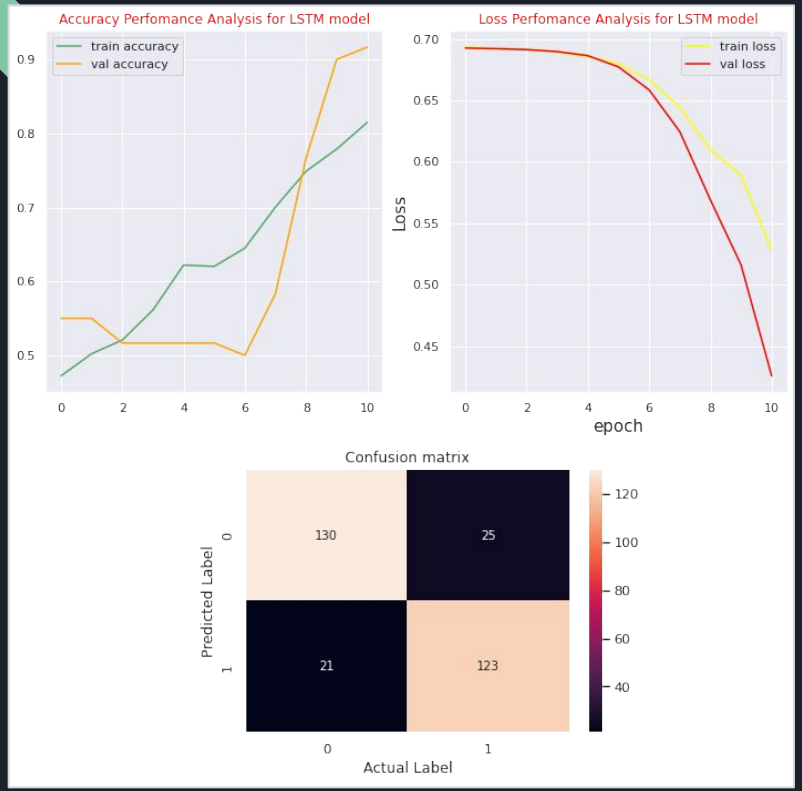


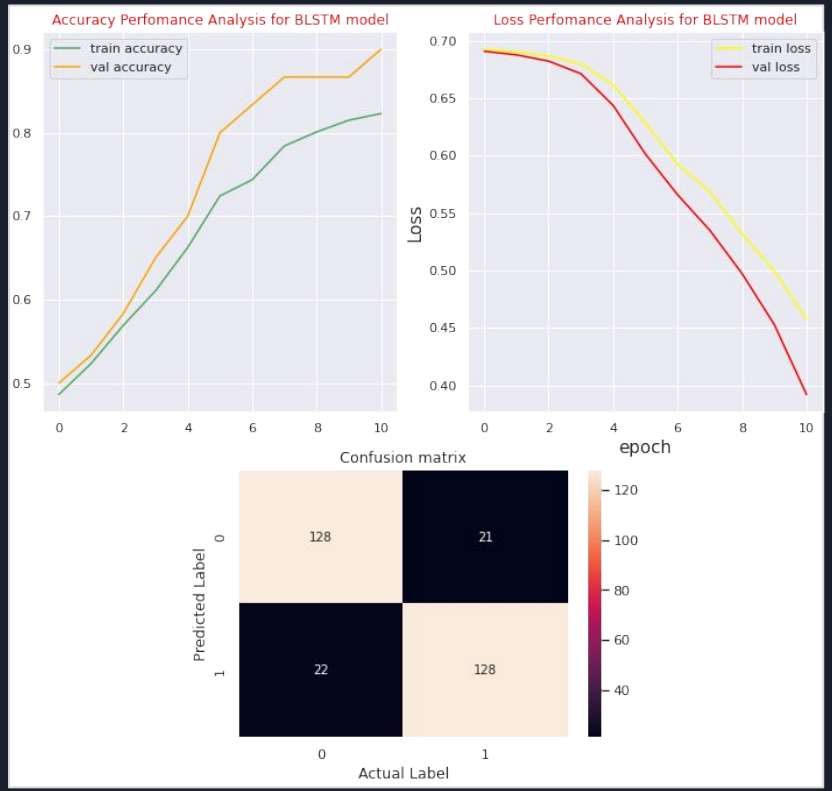
Performance Analysis - LSTM and BILSTM (Upsampling)





Performance Analysis - LSTM and BILSTM (Downsampling)





Results for models performed on upsampled data

	Accuracy	Precision `	Recall	F1-Score
Dense	95.54%	92.21%	99.37%	95.65%
LSTM	97.82%	97.27%	98.13%	97.69%
BILSTM	98.65%	98.41%	98.83%	98.62%

Results for models performed on downsampled data

	Accuracy	Precision `	Recall	F1-Score
Dense	89.97%	90.85%	89.68%	90.26%
LSTM	84.62%	83.11%	85.42%	84.25%
BILSTM	85.62%	85.91%	85.33%	85.62%

Thank you!