

Experiment

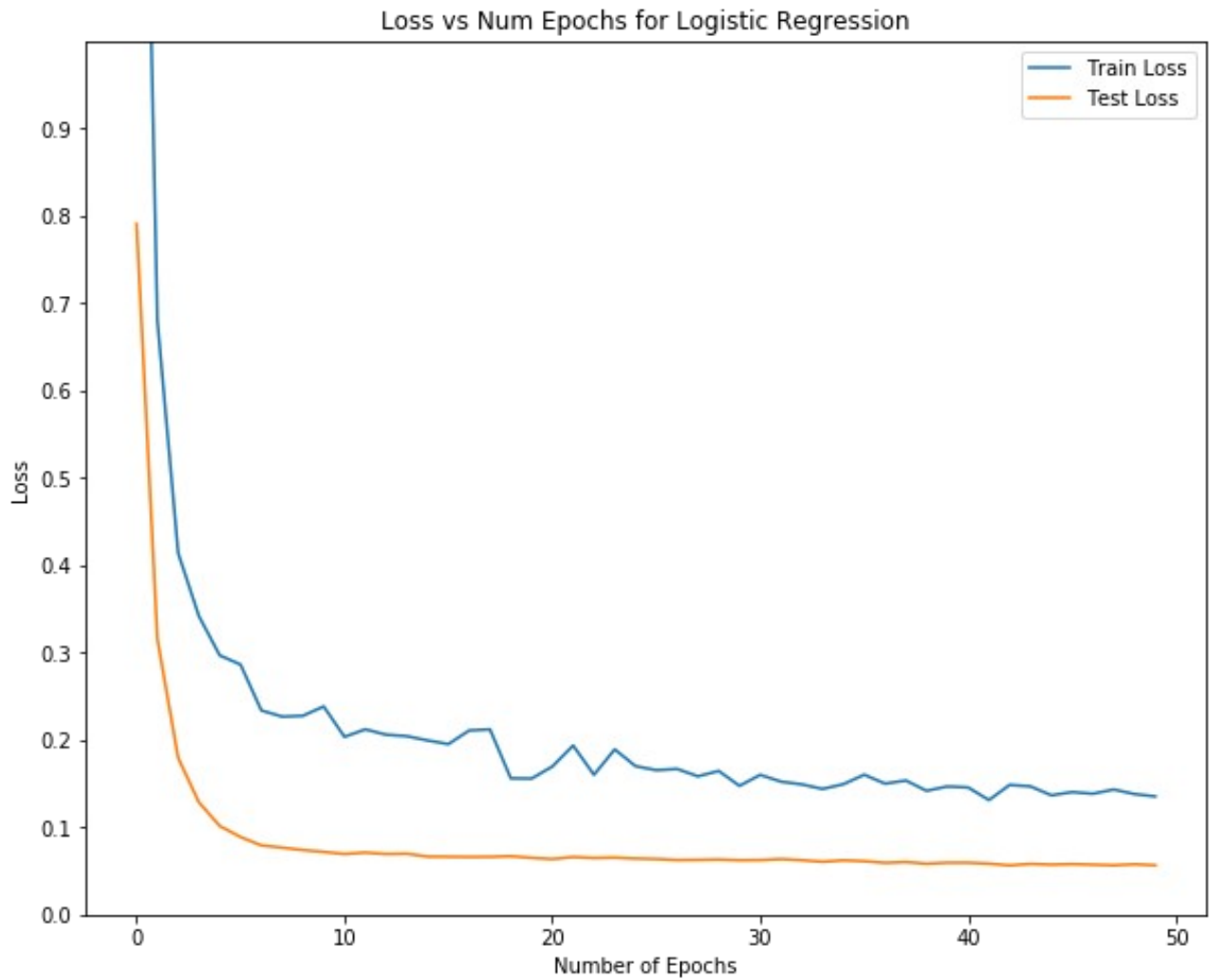
- Used loans originating in 2006-2015 as training set and loans originating in 2016-2017 as test set.
- Each epoch was trained with a subset of the training set and not the entire training set.
- After each epoch testing was done on the entire testset.
- **Since it was observed that Feed forward takes longer to converge, 50 epochs were run for Feed forward, LSTM and Logistic regression to reduce cross entropy loss.**
- AUC was calculated for transition to Prepayment status and transition to 30 days delinquency status.

Following features are used for training the LSTM, Logistic Regression, Feed forward network.

- 1) ORIGINAL COMBINED LOAN-TO-VALUE
- 3) ORIGINAL DEBT-TO-INCOME (DTI) RATIO
- 4) ORIGINAL LOAN-TO-VALUE (LTV)
- 8) CREDIT SCORE
- 9) CURRENT STATUS
- 11) CURRENT INTEREST RATE
- 14) FIRST TIME HOMEBUYER FLAG - one hot encoded
- 16) LOAN AGE
- 26) NUMBER OF BORROWERS
- 27) NUMBER OF UNITS
- 28) OCCUPANCY STATUS - one hot encoded
- 29) ORIGINAL INTEREST RATE
- 30) PRODUCT TYPE - one hot encoded
- 31) LOAN PURPOSE - one hot encoded
- 32) STATE - one hot encoded

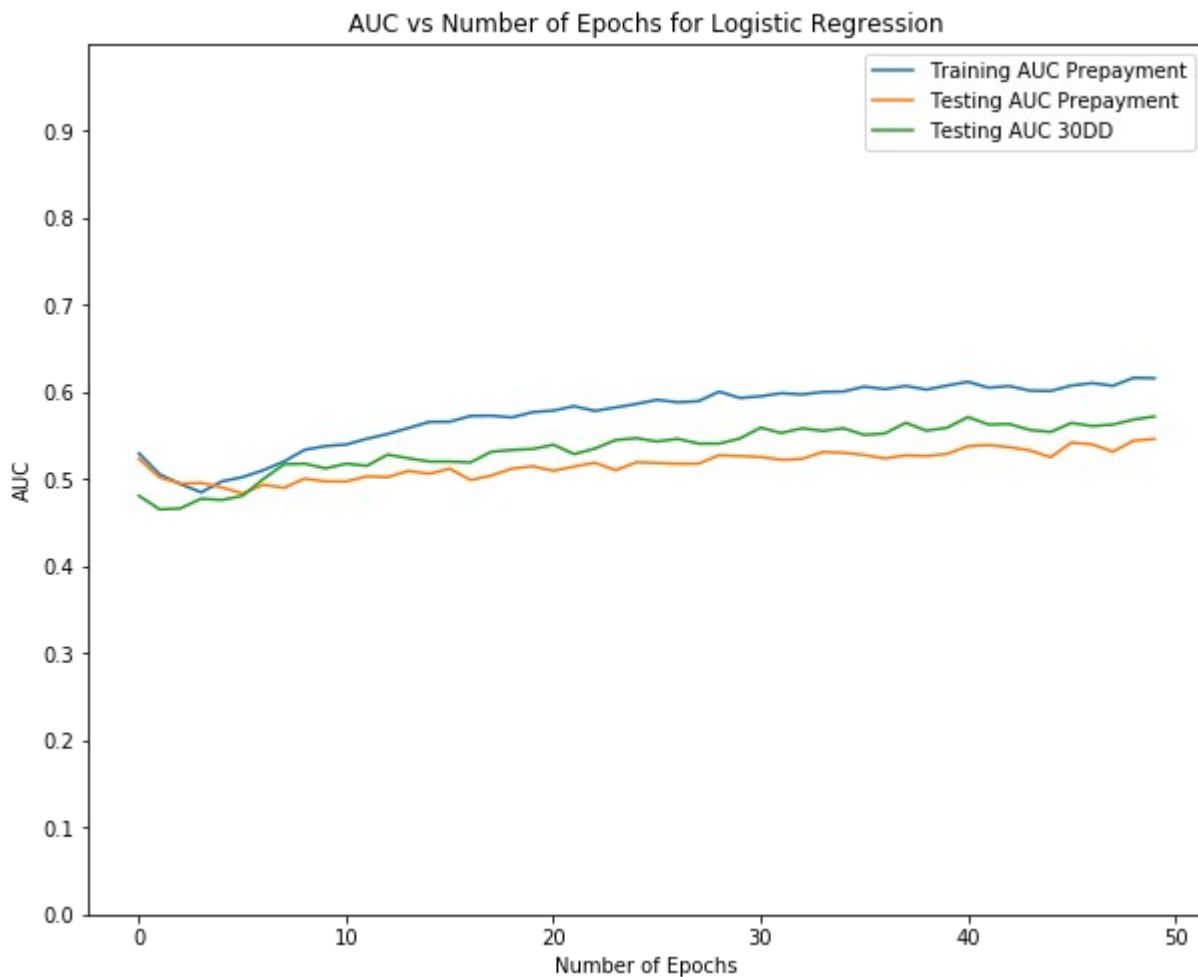
Logistic Regression

Plot of training and testing cross entropy loss against the number of epochs.



Plot of AUC for two types of transition:

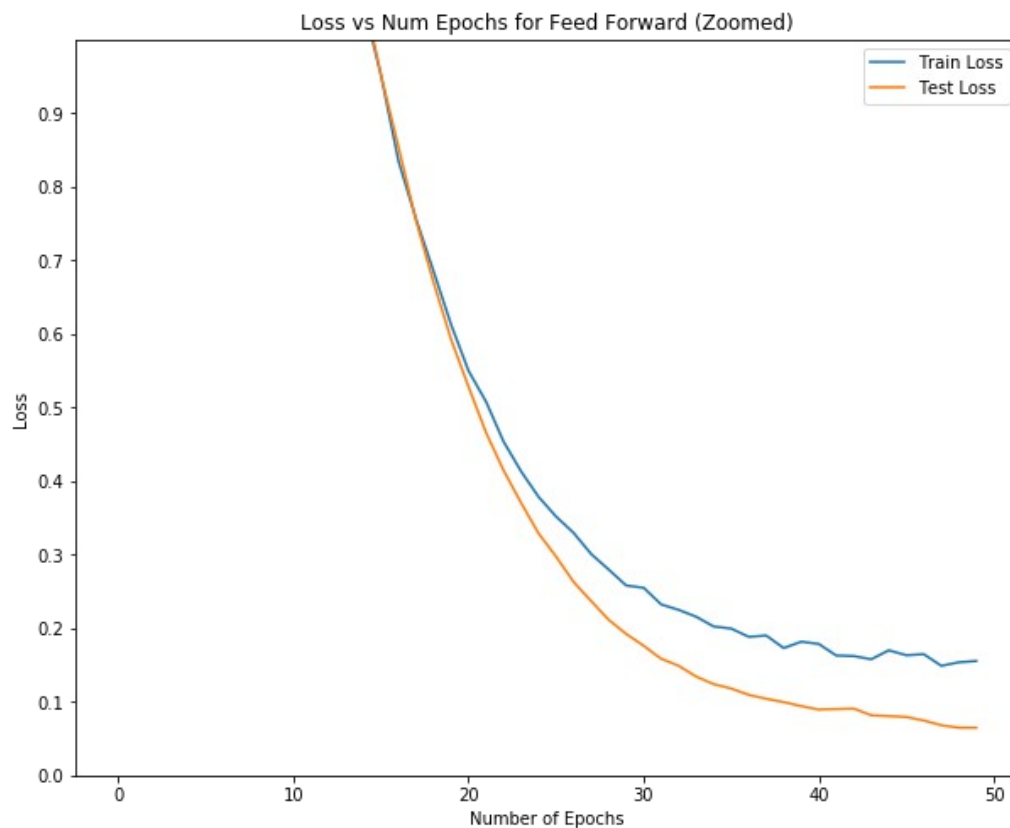
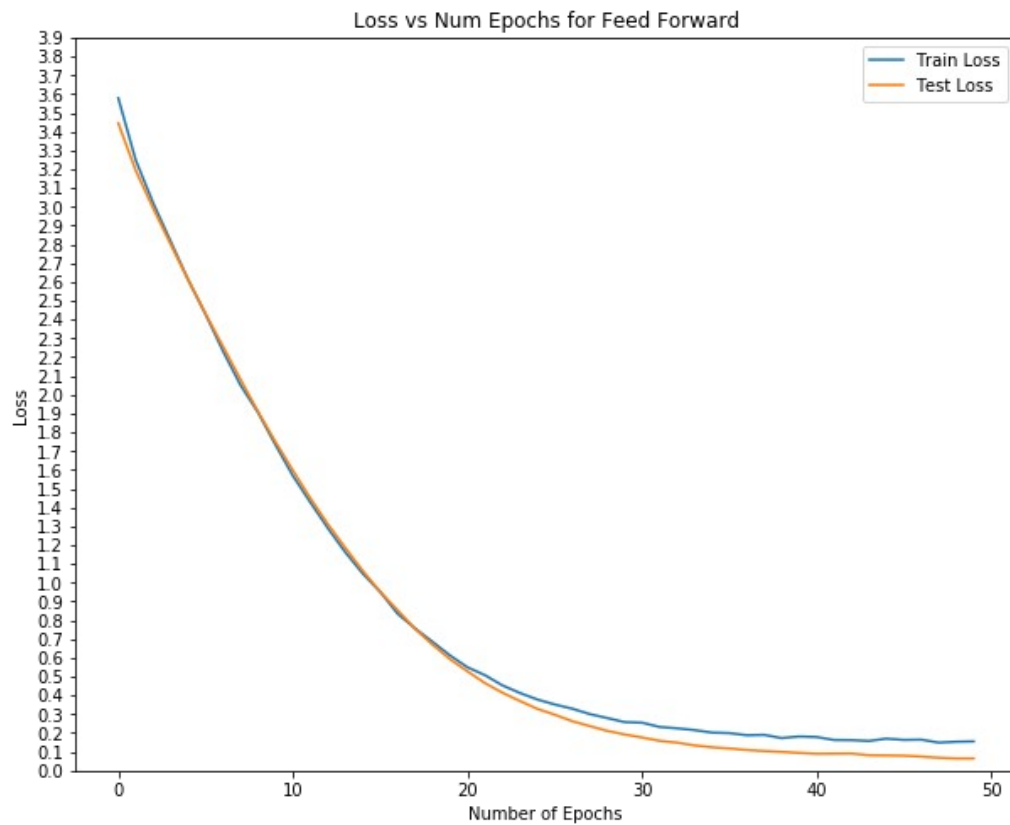
- Orange line is test AUC for transition to Prepayment status
- Green line is test AUC for transition to 30 Days Delinquency status



AUC for transition to Prepayment as well as for 30DD is around 0.5 but AUC for 30DD performs better.

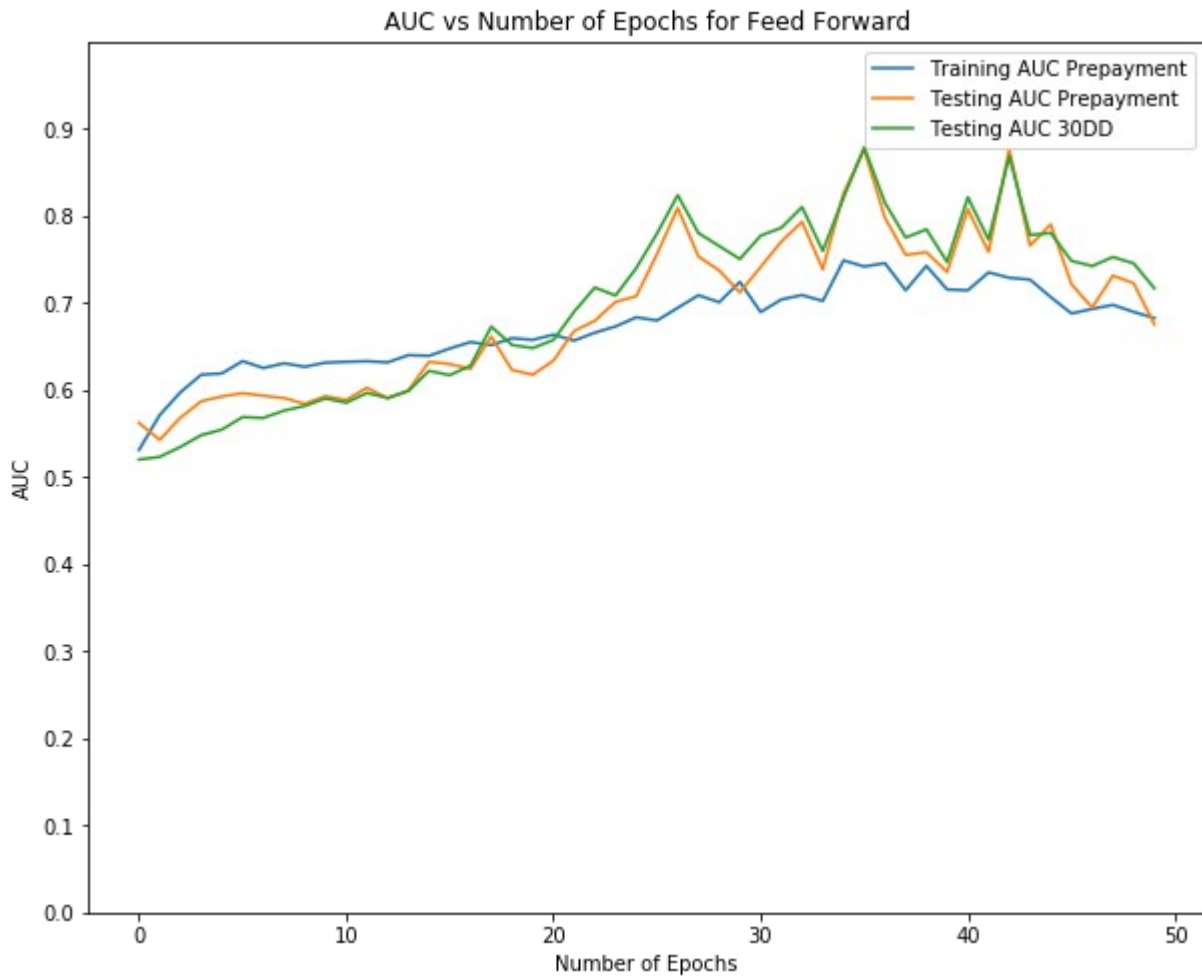
Feed Forward network

Plot of training and testing cross entropy loss against the number of epochs.



Plot of AUC for two types of transition:

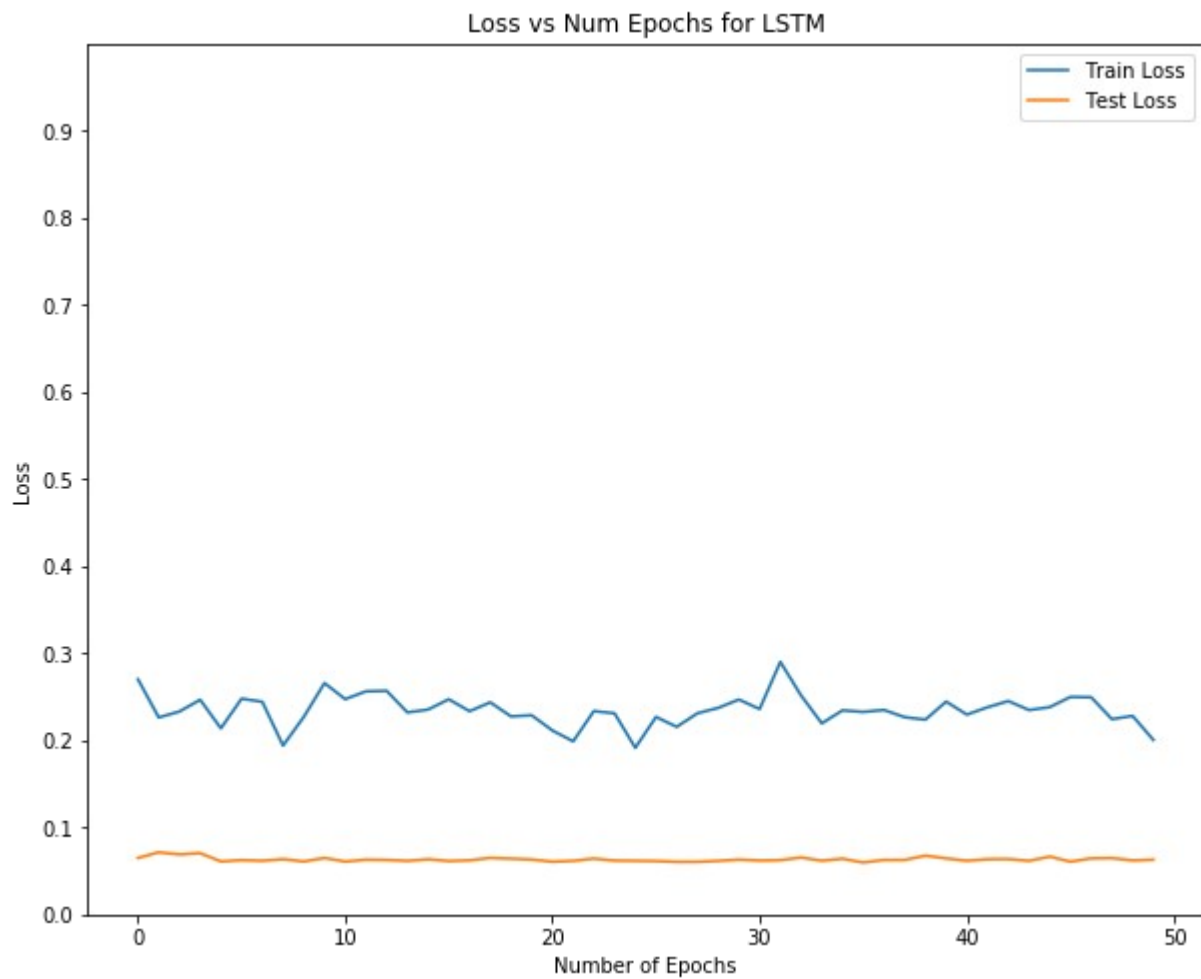
- Orange line is test AUC for transition to Prepayment status
- Green line is test AUC for transition to 30 Days Delinquency status



AUC for transition to Prepayment is around 0.7 which is an improvement on the logistic regression results but AUC for 30DD is around 0.7 here as well. With some jumps the maximum AUC is in 0.8 region.

LSTM

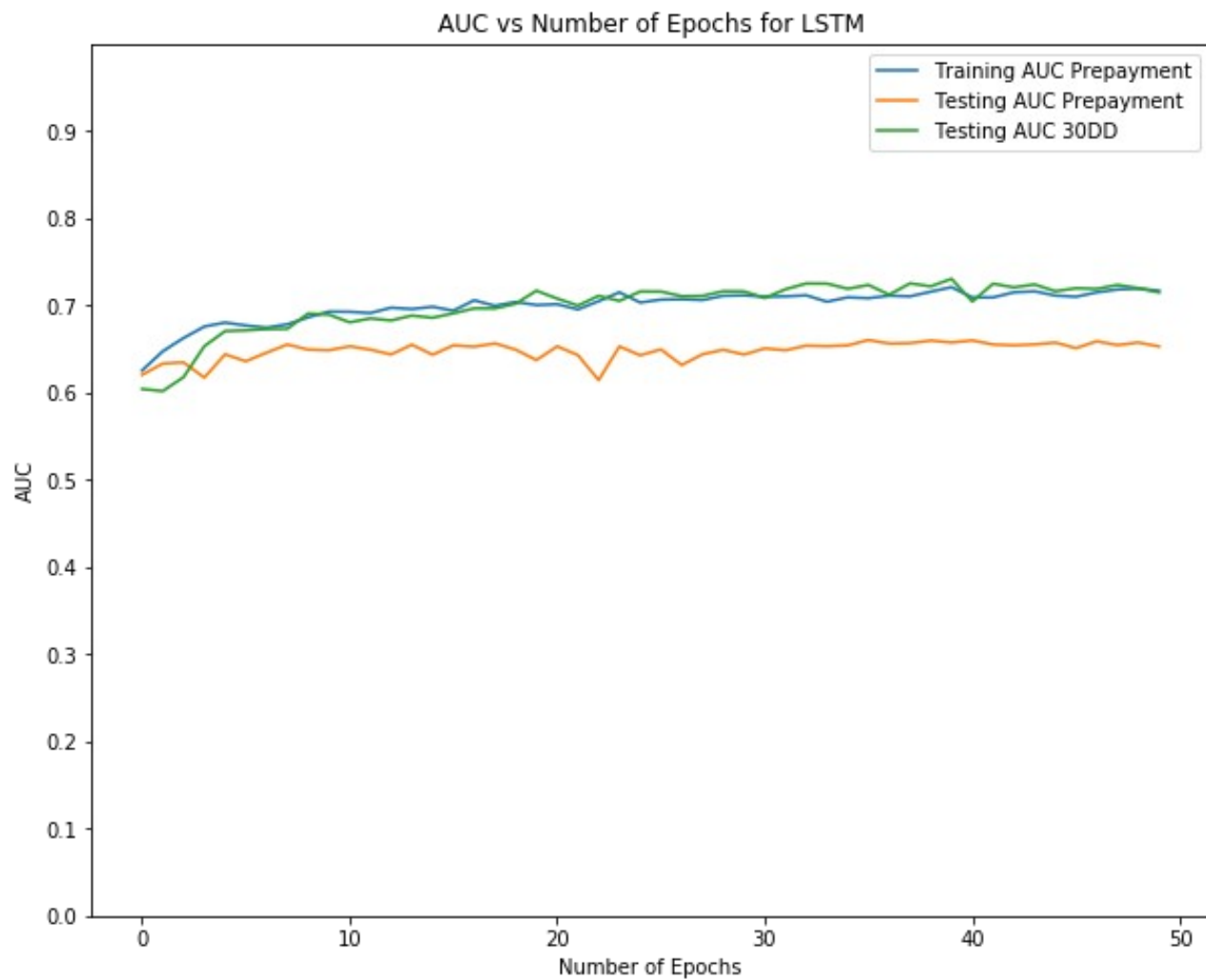
Plot of training and testing cross entropy loss against the number of epochs.



The convergence of LSTM network is extremely fast, although training time is quite low.

Plot of AUC for two types of transition:

- Orange line is test AUC for transition to Prepayment status
- Green line is test AUC for transition to 30 Days Delinquency status

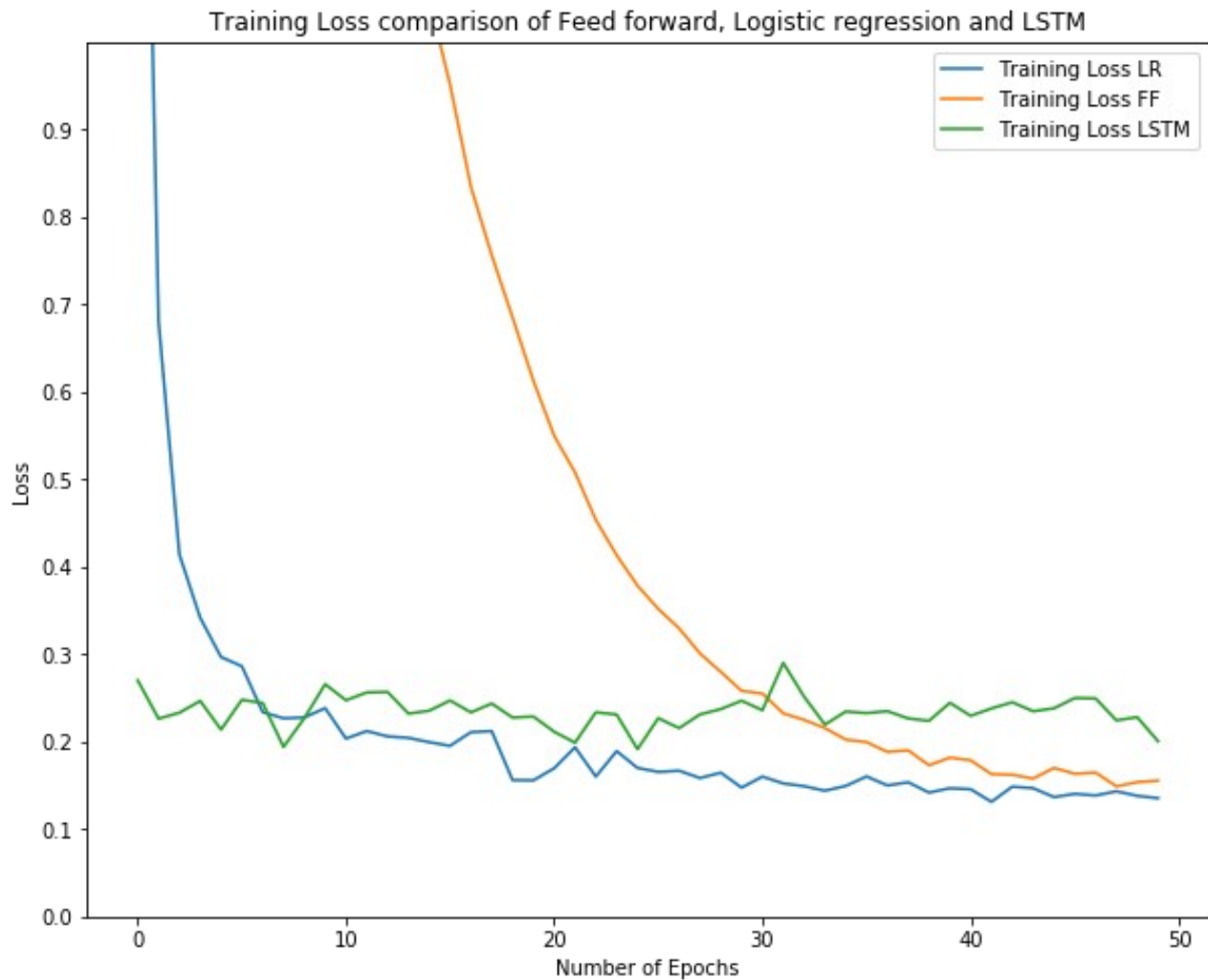


Here also we see that AUC for prepayment is a little above 0.6 and AUC for 30DD is around 0.7.

Now lets see the comparative study of the three models.

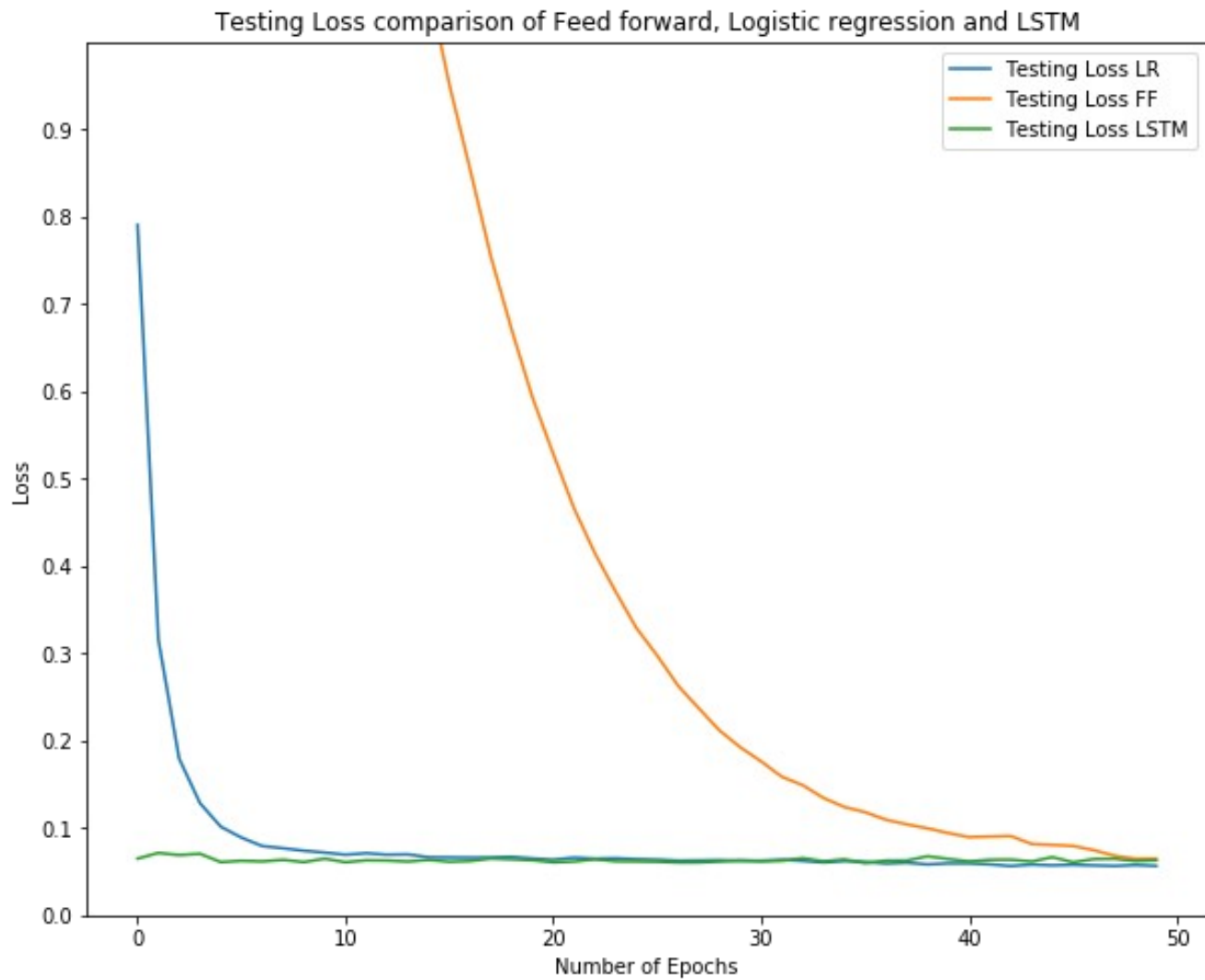
Comparison of Logistic regression, Feed forward network and LSTM

Plot of training cross entropy loss against the number of epochs.



- Feed forward takes time to converge.
- LSTM loss is quite smooth and does not fluctuate much, so it might be easier for the LSTM to converge.

Plot of testing cross entropy loss against the number of epochs.



Again FF takes longer to converge but it also converges to similar values of test loss.

Plot of AUC of transition to Prepayment on test set.

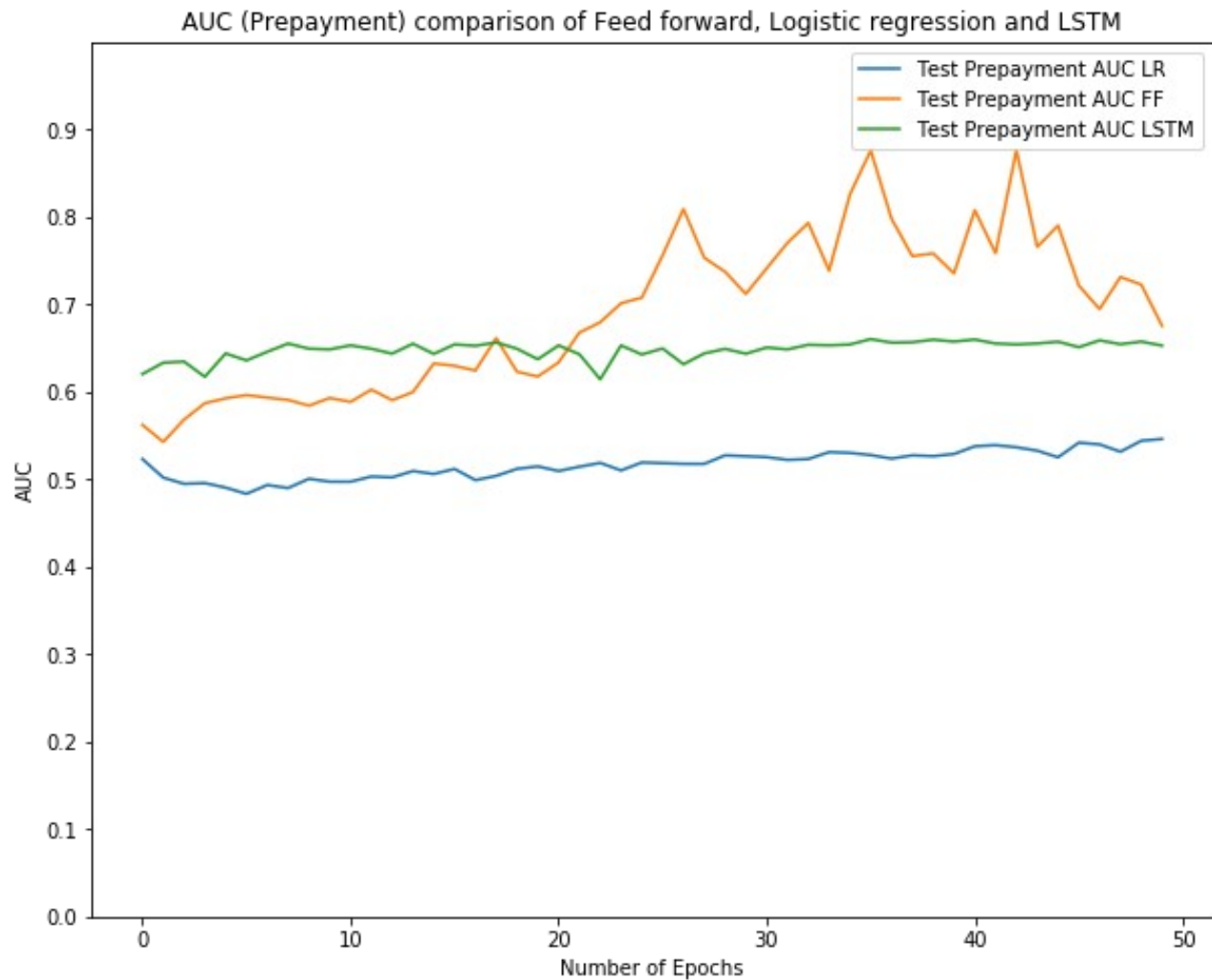


Table to show Max AUC and Mean AUC for Prepayment delinquency.

	Max AUC	Mean of AUC
Logistic Regression	0.546249	0.517159
Feed forward network	0.876679	0.689631
LSTM	0.660445	0.648359

Logistic Regression performs much worse than Feed forward network or LSTM network. Feed forward as expected performs better (on average) than LSTM and logistic regression.

Plot of AUC of transition to 30DD test set.

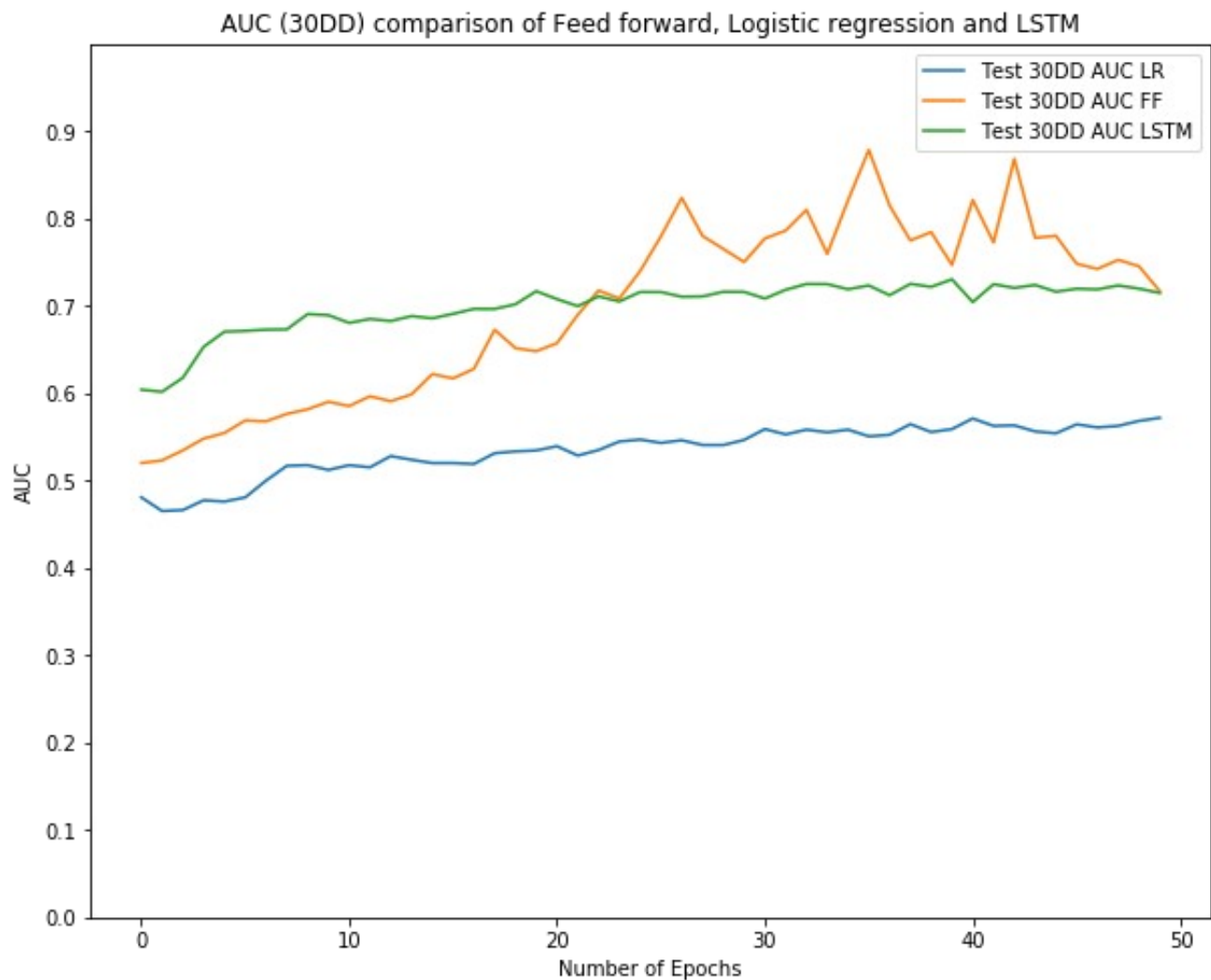


Table to show Max AUC and Mean AUC for 30 days delinquency.

	Max AUC	Mean of AUC
Logistic Regression	0.572034	0.535181
Feed forward network	0.878842	0.697579
LSTM	0.730657	0.699703

LSTM and FF work similar on average and better than Logistic regression. But the highest AUC is that of FF.