Flight Delays Predictions

Lara Abu Hamad
Elona Mendes
Firas Naamne

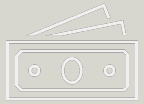
Introduction

This project to predict flight delays dataset from Bureau of transportation.

Flow

Data analysis

Clean, scale, split data



Models

Softmax, multiple hidden model, ANN model.

Results and Comparison

Accuracy, confusion matrix.

Data analysis

Unique

Check unique variables.

Clean

Drop columns only have 1 unique values and useless columns.

Scale

Standard scale data

SMOTE

Balanced data

Models

Softmax

1 softmax layer

3 hidden models

Only 3 hidden layers.

5 hidden model

Only 5 hidden layers.

ANN model

6 hidden layers.

ANN Architecture

Dense layer

Deeply connected with its preceding layer which means the neurons of the layer are connected to every neuron of its preceding layer.

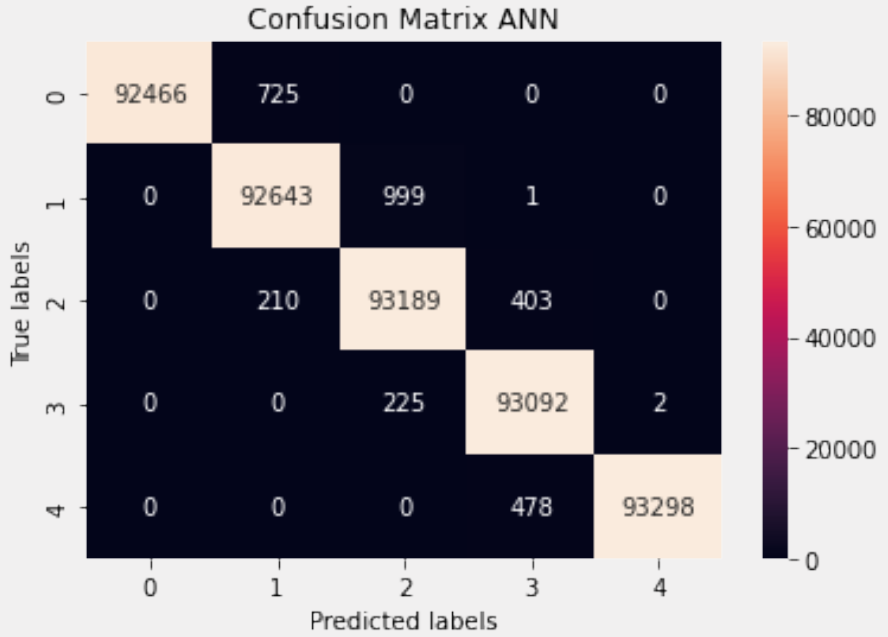
Dropout layer

Randomly selected neurons are ignored during training. They are “dropped out” randomly. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass, and any weight updates are not applied to the neuron on the backward pass to avoid overfitting.

Result.

Model	ANN
Accuracy	0.99
Precision	0.99
Recall	0.99
F1-score	0.99

Result.


Confusion Matrix ANN
True labels
Predicted labels
0
1
2
3
4
92466
725
0
0
0
0
92643
999
1
0
0
210
93189
403
0
0
0
225
93092
2
0
0
0
478
93298
80000
60000
40000
20000
0

Result.

Report model ANN

	precision	recall	f1-score	support
0	0.99	1.00	1.00	92466
1	0.99	0.99	0.99	93578
2	0.99	0.99	0.99	94413
3	1.00	0.99	0.99	93974
4	0.99	1.00	1.00	93300
accuracy			0.99	467731
macro avg	0.99	0.99	0.99	467731
weighted avg	0.99	0.99	0.99	467731

Comparison

Models	Softmax	3 hidden layers model	5 hidden layers model	ANN
Num. of layers	1	3	5	7
Optimizer	GradientDescentOptimizer	Adam	GradientDescentOptimizer	Adam
Loss	log	softmax_cross_entropy_with_logits_v2	softmax_cross_entropy_with_logits_v2	Categorical cross entropy
Batch size	300	300	300	128
Epochs	300	300	1000	20

Comparison

Models	Softmax	3 hidden layers model	5 hidden layers model	ANN
Accuracy	0.65	0.92	0.75	0.99

Conclusion

We do data analysis, check balance dataset, deploy models to achieved perfect fits on the testing set. Fully achieved both of our research objectives. We developed ANN model to detect flight delay and accuracy is 0.99.

There are several directions for improvement can be taken. The Fisher's score or Pearson's Correlation Coefficient. Or use another better model like LSTM,…

Thank you

Lara Abu Hamad

Elona Mendes

Firas Naamne