Abstract: Accurate flight departure time prediction enables the rational use of airport support resources, aprons, and runway resources, and promotes the implementation of collaborative decision making. In order to accurately predict the flight departure time,we will apply deep learning. First, we analyzes the influence of different factors on flight departure time and the influencing factor then clean the data. Secondly, we build model from tensorflow library, considers the impact of different hyperparameters on network performance, and determines the optimal hyperparameter combination through parameter tuning.

This paper presents a process of improving results using different ML methods.
First, we used Softmax with test accuracy of 66%.
Second, a model with 3 hidden layers with test accuracy of 92%.
Finally, we used a model with 5 hidden layers with test accuracy of 77%.

We improved the accuracy to 99% using ANN model with 6 hidden layers of DROPUT and DENSE.

All models were trained and tested on the same dataset that contains 24 features.

1. Introduction:

People and companies today are connected around the world, which has led to a growing importance of the aviation industry. As flight delays are a big challenge in aviation. When affected by delays, dissatisfaction grows among passengers because of missed connec-tion flights. Both airlines and airports have to deal with increased costs and challenges in planning. Even though there is an ongoing effort in improving the current air traffic management processes to minimise delays. This project to predict flight delays dataset from Bureau of transportation.

2. Related work and required background:

Most studies work as follows: firstly, they analyzed and screened the factors that may cause flight delays. Then, they built a delay prediction model and finally used machine learning or deep learning algorithms to solve the problem but they use too much variables to build model. Departure time prediction does not need to consider many factors, and the specific flights can be focused on. Therefore, a flight departure time prediction model is proposed to screen the influencing factors, reduce the modeling complexity, and finally obtain an accurate flight departure time, which can provide a reliable basis for airport scheduling and the implementation of collaborative decision-making systems.

First, we summarizes the factors affecting flight operation in existing research results, and analyzes and filters the factors, so as to determine the factors affecting flight operation.

Then, the ANN neural network model is established, which is verified by the flight data.

Finally, compared with commonly used neural network models.

Relevant Links:

1) https://www.mdpi.com/2071-1050/14/22/15367

2) https://medium.com/analytics-vidhya/using-machine-learning-to-predict-flight-delays-e8a50b0bb64c

3) https://industryforever.wordpress.com/2017/10/12/predicting-flight-delays-using-tensorflow-and-machine-learning/

3. Project description:

After read data by pandas we will check null values in the dataset and we need to remove the null values. It is important to handle the missing values appropriately.

	YEAR	QUARTER	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	FL_DATE	MKT_UNIQUE_CARRIER	BRANDED_CODE_SHARE	MKT_CARRIER_AIRLINE_ID
0	2022	3	8	18	4	8/18/2022 12:00:00 AM	DL	DL_CODESHARE	19790
1	2022	3	8	18	4	8/18/2022 12:00:00 AM	DL	DL_CODESHARE	19790
2	2022	3	8	18	4	8/18/2022 12:00:00 AM	DL	DL_CODESHARE	19790
3	2022	3	8	18	4	8/18/2022 12:00:00 AM	DL	DL_CODESHARE	19790
4	2022	3	8	18	4	8/18/2022 12:00:00 AM	DL	DL_CODESHARE	19790
...	...	...	...	...	...	...	...	...	...
613644	2022	3	8	1	1	8/1/2022 12:00:00 AM	UA	UA_CODESHARE	19977

Some variables on have 1 unique value, so that we need to remove these column to clean our data. And then only choose which variables relate on and helpful to predict flight delay.

We will drop these variables: 'YEAR', 'QUARTER', 'MONTH',… because it's datetime so that can't help on delay result. The processed data final has 24 columns. Then replace null value with 0 for all 24 columns.

Out[5]:

	DAY_OF_MONTH	DAY_OF_WEEK	MKT_CARRIER_AIRLINE_ID	MKT_CARRIER_FL_NUM	ORIGIN_AIRPORT_ID	DEST_AIRPORT_ID	CRS_DEP_TIME	DEP_DELA
0	18	4	19790	4036	12197	10397	1240	
1	18	4	19790	4037	11433	10408	2146	
2	18	4	19790	4038	11433	10408	1555	
3	18	4	19790	4039	10408	11433	1755	
4	18	4	19790	4040	11898	13487	1311	

5 rows × 24 columns

We want to merge 5 columns, giving each reason a number 1-5:

- 'CARRIER_DELAY'
- 'WEATHER_DELAY'
- 'NAS_DELAY'
- 'SECURITY_DELAY'
- 'LATE_AIRCRAFT_DELAY'
  into one column **'DELAY_REASON'**.
  0 defines the flights that didn't get delayed

And then Dropping the columns:

- 'CARRIER_DELAY'

- 'WEATHER_DELAY'
- 'NAS_DELAY'
- 'SECURITY_DELAY'
- 'LATE_AIRCRAFT_DELAY'

Prepare the label 'ARR_DELAY' we want to predict.

- Less than 15 minutes: 0
- 15 minutes to 30 minutes: 1
- 30 minutes to 45 minutes: 2
- 45 minutes to 60 minutes: 3
- more than 60 minutes: 4

So we have total 5 classes need to predict. Because of imbalanced data so that next step we will apply SMOTE to random generate data to have balanced dataset. It is an Oversampling technique that allows us to generate synthetic samples for our minority categories.

We split data to train/val/test: 0.6, 0.2, 0.2. Total 596284 rows of data.

- Model ANN:

n	1	2	3	4	5	6	7
Layers	Dense 512	Dense 256	Dense 128	Dropout 0.3	Dense 64	Dropout 0.2	Dense 5 (softmax)

Dense layer is a layer that is deeply connected with its preceding layer which means the neurons of the layer are connected to every neuron of its preceding layer.

Dropout is 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.

All dense layers we use relu activation to active learning curve. Besides we use Adam optimizer function and categorical cross entropy as loss function.
Then we train 20 epochs to get best accuracy of model.

4. Results:

Evaluation is very important in data science. It helps you to understand the performance of your model. There are many different evaluation metrics out there but only some of them are suitable to be used for classification.

In this project we will use 3 metrics to evaluate our models: Accuracy, Confusion matrix and classification report (precision, recall, F1-score).

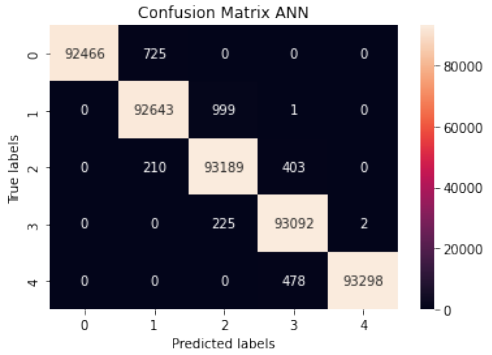
y_i: True value

$\tilde{y}_i$: Predicted value

Model	ANN
Accuracy	0.99

Precision	0.99
Recall	0.99
F1-score	0.99

Confusion matrix:



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

From the confusion matrix we can see model performed very well, just a little wrong classes 0 and class 1.

5. Comparison:

This section we compare 4 models: softmax, 2 different hidden models and ANN model.

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

Models	Softmax	3 hidden layers model	5 hidden layers model	ANN
Accuracy	0.66	0.92	0.77	0.99

⇨ The highest accuracy of ANN model: 0.99 when the worse model is Softmax model. Because ANN model has 6 hidden layers also apply dropout to avoid overfitting so that model can archive best accuracy.

6. Conclusions:

In this project, we did some research models and data analysis.

First, computational methods are used to quantitatively determine the importance of each feature. We do data analysis, check balance dataset, deploy models to achieved perfect fits on the testing set. This led us to conclude that standard deep learning models are sufficient to act. Through or methodology, data collection, and analysis, we 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. One of them is feature selection can be used such as the Fisher's score or Pearson's Correlation Coefficient in order to validate our results. Or use another better model like LSTM,…