

# Study of Prediction Algorithms on Aviation Accident Dataset using Rapid Miner

Dishank Kaji<sup>1</sup>, Dhrumil Mehta<sup>2</sup>, Divyesh Sanghani<sup>3</sup>, Harsh Shah<sup>4</sup>, Rashmi Malvankar<sup>5</sup>

<sup>5</sup>Under the Guidance of Prof. Rashmi Malvankar, Department of Information Technology,  
<sup>1,2,3,4</sup>Shah and Anchor Kutchhi Engineering College, Mumbai, Maharashtra, India.

\*\*\*

**Abstract** - A Flight crash is caused due to multiple factors which can lead to end of many life. In the field of aircraft, Person's safety is always considered as an important factor. Further Analyzing any kind of an incident in aircraft will help to save life and cost. Prediction is made based on huge collection of flight records using data mining techniques and algorithms. To minimize these accidents, factors causing or contributing to accidents must be prevented. Flight crash can be caused due to n number of factors. In this research work, comparative analysis of the different algorithms like Decision tree and Naive Bayes will be implemented and a decision will be made on various parameters whether a flight is ready to take off or not. The analysis identifies relationships between the accident and incident data and finds patterns of contributory factors which are significantly associated with aircraft accidents.

**Key Words:** Decision Tree, Naive Bayesian

## 1. INTRODUCTION

Data mining has been intensively and extensively used in many fields and applications. The data generation and collection are considered as an important factor for producing data sets from different kind of sources. Data is comprehensively collected from NTSB, which records all the aircraft accidents and it is used as training dataset for the proposed system. Various attributes which caused the accidents are analyzed. Huge collections of past accident records were analyzed from different reports and record were filtered accordingly to perform different algorithms and each record was found to be erroneous. Decision Tree is implemented in a more effective way to handle nonlinear data. The Naive Bayesian classifier is based on Bayes' theorem and it is executed in an effective way such that it compares records with the independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly for huge collection of data.

## 2. PROBLEM DEFINATION

To work with unstructured data and processing it dynamically is very difficult, requires effort, man power and is highly time consuming and there is high percentage of chances that low standard results will be obtained. Airline Accident is caused by numerous factors and data of flight crash are unpredictable. So, to process huge

collection of data that has the tendency to change frequently, various automated tools must be used. In this case Rapid Miner.

## 3. DATA COLLECTION

Data is collected from National Transport Safety Board (NTSB). NTSB keeps record of all the accidents occurred across all means of transport and has been doing so for more than ten years. We will be using their aviation accident dataset which has more than 70,000 records which gives more data to algorithms to train. This data is available in .xml, .csv and .txt format.

## 4. DATA SELECTION AND PREPROCESSING

The NTSB dataset consists of data with large number of attributes like Type of Event, Location, Longitude, Latitude, Injury Severity, Aircraft Category, Make, Engine type, Schedule, Purpose of Flight, Total fatal injuries, Weather Conditions, Total Uninjured.

Not all of these attributes contribute towards the crash and so the data must be filtered i.e. only the relevant attributes must be chosen rest all shall be excluded and smoothen i.e. the missing values must be taken care of. We have replaced the missing value with the mean. The attributes chosen that are relevant towards predicting the crash are:

- i) InjurySeverity: Highest level of injury.
- ii) AircraftCategory: It defines the rating, certification, limitations and privileges of an air personnel.
- iii) Make: Build of the aircraft.
- iv) AmateurBuilt
- v) Weather condition: There are two types of weather conditions when concerned to aviation. First IMC (Instrument meteorological conditions) and second VMC (visual meteorological conditions)

### 3. ALGORITHMS OF PREDICTION

#### A. DECISION TREE

Decision Tree algorithm is a member of supervised learning algorithms. Decision tree algorithm is helpful in filtering regression and classification problems. The objective of this algorithm is to create a training model which can be used to predict aircraft accident or decision rules inferred from prior data. First the whole training set is considered as the root. Then Feature values are preferred to be categorical. If the values are continuous then they are discretized prior to building the model. Records are distributed recursively on the basis of attribute values. Order to placing attributes as root or internal node of the tree is done by using some statistical approach. ID3 is the decision algorithm used for accident prediction. ID3 uses information gain as its attribute selection measure. It decides which attribute (Splitting point) to test at node N by determining the best way to separate or partition the tuples into individual classes.

because it takes enormous numbers of observations to estimate the probabilities. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform classification methods.

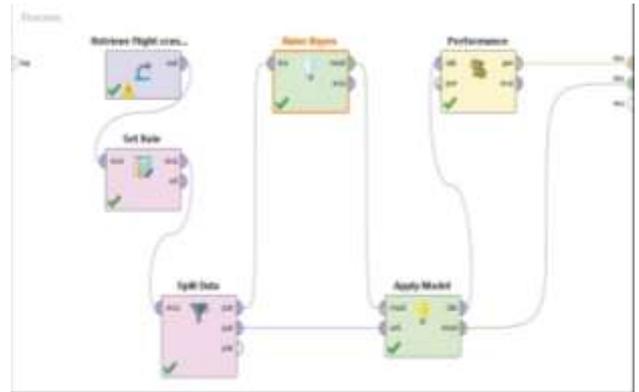


Fig-2: Naive Based (Rapid miner tool)

RESULT: The Accuracy obtained was 97.31%

PERFORMANCE VECTOR:

Table-2: Performance Vector of Naive Bayesian

Parameters	True Yes	True No	Class Recall
Pred Yes	22011	456	97.97%
Pred No	180	1015	84.94%
Class Recall	99.19%	69.00%	

#### 4. CONCLUSION

Data mining platform is very helpful in performing prediction analysis of an improper dynamic data. Prediction of aircraft accident is a critical factor. Thus, it has been implemented with two efficient data mining algorithms Naive Bayes and Decision Tree. These algorithms are efficiently used in many fields. Two algorithms act efficiently in different real time conditions. On usage of these algorithms' aircraft accident has been predicted with standard accuracy.

#### 5. REFERENCES

[1] [https://www.ripublication.com/ijaer18/ijaerv13n21\\_01.pdf](https://www.ripublication.com/ijaer18/ijaerv13n21_01.pdf).  
 [2] <http://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/>.  
 [3] <https://ieeexplore.ieee.org/document/6528602>.

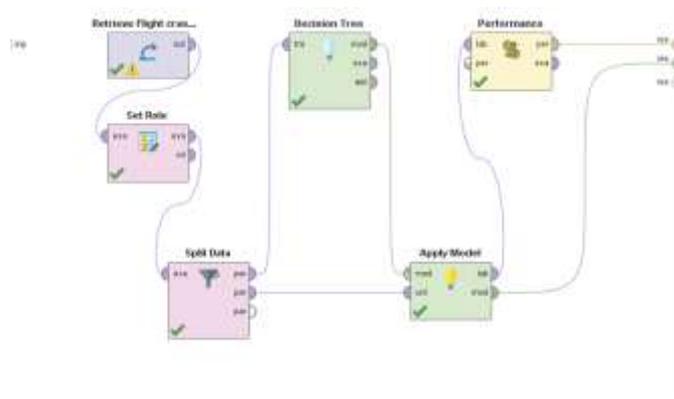


Fig-1: Decision Tree (Rapid miner tool)

RESULT: The Accuracy obtained was 97.77%

PERFORMANCE VECTOR:

Table-1: Performance Vector of Decision Tree

Parameters	True Yes	True No	Class Recall
Pred Yes	22152	488	97.84%
Pred No	39	983	96.18%
Class Recall	99.82%	66.83%	

#### B. NAÏVE BAYESIAN

A Naive Bayes (NB) classifier is a simple probabilistic classifier based on Bayes theorem where every feature is assumed to be class-conditionally independent. In naïve Bayes learning, each instance is described by a set of features and takes a class value from a predefined set of values. Classification of instances gets difficult when the dataset contains a large number of features and classes

[4]<http://dataillumination.blogspot.com/2015/03/predicting-flights-delay-using.html>

[5][https://www.ripublication.com/acst17/acstv10n3\\_10.pdf](https://www.ripublication.com/acst17/acstv10n3_10.pdf)

[6][https://www.researchgate.net/publication/321362053\\_Accident\\_Analysis\\_by\\_using\\_Data\\_Mining\\_Techniques](https://www.researchgate.net/publication/321362053_Accident_Analysis_by_using_Data_Mining_Techniques)

[7] <https://www.grin.com/document/386836>