Predicting Causes of Airplane Crashes using Machine Learning Algorithms

Ved Prakash Gupta, M Sajid Mansoori, Jitendra Shreemali, Payal Paliwal

Abstract: ***Considering the immense cost of air crashes, the study examines the causes of crashes of aircrafts based on reported findings for the crash. The dataset used for this study included data for all reported air crashes across the globe for the period from 1981 to 2019. The causes were classified into seven categories. Multiple machine learning algorithms were used to identify the best for predicting the likely cause of accident based on features available. The Machine Learning Models used are Auto Classifier, Tree-AS and XGBoost. Also the key predictors are identified for use by planners.***

Keywords: Machine Learning, XGBoost, Neural Network, Deep Learning, Tree-AS

I. INTRODUCTION

Airplanes are being used as a mode of transportation by millions of people on a daily basis. The invention of airplanes has done a lot of good as it is used by people for a plethora of purposes be it for travelling to exotic places , for business meetings , or just to meet with one's family members or friends. The airplanes are getting popular as a means of transportation amongst people by the day mostly due to the fact that they are the fastest mode of transportation available. Also, people's jobs nowadays are pretty tedious and demanding and time efficiency is a very essential factor to stay competitive among one's colleagues. Apart from this, aviation has proved to be the safest mode of transportation. Generally speaking, a person is more likely to die in a car accident than in a plane accident because the rate of aviation accidents is much lesser than car accidents. However, it is true that the chances of surviving a car accident are much better than in a plane accident as the results of an airplane crash may result in catastrophic numbers of death. Since the first flight by the Wright Brothers there has been a substantial improvement in the technologies and machineries which have contributed to making the planes a lot safer than they used to be.

Even though airplanes have their own advantages, they do come with their fair share of disadvantages and problems. One of the cons that airplanes have is that they consume enormous amounts of fossil fuel which pose a great threat to environmental reserves of fossil fuels. In addition to this, the airplanes when burn large amounts of fuel also play a major role in polluting the environment.

Also, airplanes are greatly affected by environmental conditions. It gets particularly difficult to navigate and control an airplane in bad weather. Strong winds cause turbulence, fog causes visual hindrance as well.

Also the chances of surviving an aviation accident are almost close to zero.

The most common reasons for aviation accidents are mentioned below :

1. Mechanical Faults : Since airplanes are unequivocally very complex machines there certainly persists a chance of some equipment failing to work as expected.
   Even if there is a malfunctioning of a small component, it may lead to a chain reaction of things going wrong which eventually may lead to loss of control of the aircraft resulting in major airplane accidents.
2. Pilot & Crew Error : It is one of the most frequent causes of airplane wrecks. Airplanes are unquestionably very complex feats of engineering and flying them is not a simple task either. It requires a great amount of precision and concentration while operating an airplane. There are a lot of gauges as well as handouts that have to be read and understood. Even a minor error while reading or understanding them can lead to major failures resulting in fatal crashes. So a pilot has to work in an extreme amount of pressure with utmost sincerity.
3. Environmental Conditions : Environmental conditions certainly play a role in airplane accidents. Since an airplane flies at very high altitudes if the weather deteriorates it becomes very difficult to control the airplane. If there's fog in the environment it becomes difficult to navigate mid air. Also, if there are gale winds in the
   environment the airplane may experience turbulence resulting in

loss of controls. It is also seen that many of the plane crashes are a result of a lightning strike on the plane resulting in major equipment going haywire. Unlike other vehicles, airplanes can't be immediately put to a halt if certain conditions are encountered as they need a suitable runway to land.

4. Air Traffic Control Errors : Sometimes there may be an accident due to the error made not by the pilot and his crew but instead by the air traffic controller communicating with them. It is not unheard of that an air traffic controller was overworked and dozed off at the time of duty or he/she made an error while helping the pilots navigate their way back to the airports resulting in airplane crashes.
5. Other Factors : Apart from the above mentioned reasons there are some other factors that have contributed to airplane crashes in the past. One of these reasons is war. During war, the planes have been shot down by ground missiles causing them to crash.

Revised Manuscript Received on March 5, 2020.

Ved Prakash Gupta*, Techno India NJR Institute of Technology. vedprakashgupta2309@gmail.com
M Sajid Mansoori, Techno India NJR Institute of Technology.
Jitendra Shreemali, Techno India NJR Institute of Technology.
PayalPaliwal, Techno India NJR Institute of Technology.

Also, if a plane has accidentally flown into anti-aircraft zones they are shot at. Apart from these factors, it's seen that terrorist attacks and hijacking has also played a role in airplane crashes such as the attack on the world trade center, USA. Sometimes, airplanes are sabotaged by terrorist organizations as well. In addition to all these reasons , some other reasons although very rare are the pilots fainting or dying on the airplane due to medical reasons , pilots being mentally ill trying to crash the airplanes, some passengers trying to attempt suicide jeopardizing other lives as well.

II. LITERATURE REVIEW

A non-official organization known as Bureau of Aircraft Accidents Archives (a.k.a B3A) rooted in Geneva, Switzerland has the sole purpose to compile statistics on aviation accidents of aircrafts which are qualified to carry more than six passengers, excluding helicopters , balloons and also warfare aircrafts. According to B3A the total number of fatalities because of accidents including aircrafts is roughly around 83772 since 1970.

According to a report by the Airport Apron Management and Control Program (ACRP), the recent years i.e. , from 2009 - 2017 have been much safer with less than 170 incidents a year as opposed to the year 1998 which has the count of more than 226 incidents.

Retired Chief Pilot from United Airlines Capt. Randy DeAngelis says that the pilots should be trained more efficiently so that if the aircraft any problem such as adverse weather conditions they should be able to maintain control of the aircraft. The important point is to invest in Human Factor Research so that the risks due to human factors will be reduced significantly. (Source : http://www.arnellent.com/aviation-crew-Randy-DeAngelis.html)

Stephens andUkpere (2014) in their paper say that on examination of causes of air crashes over time has shown that during the time when airplane travel was still in infancy, most of the crashes were due to fuel starvation, some flaws in the aircraft design and lightning. Also, during those times the manufacturing methods as well as technologies used were primitive in contrast to today's.

Chen ,PHUN, YAI and SUZUKI (2015) in their research about how an aviation accident affects the point of view of the public found out that airline companies take a hit if there are multiple cases of crash. The direct result of such crashes is the airlines stock taking a dive as well as a loss of customers. So they have concluded that public perception and their trust is a very important factor in choosing an airline or even deciding whether or not to travel by air in the first place.

In addition to this, the governing body in the accident prone area plays a crucial role in how much decline stock prices observe. The researchers Walker, Thie., Walker M.. and K.P. (2014) have conducted research on this and they found out that in the US if an accident is covered and governed by the state law which doesn't put any limit over the damage claims result in quite a huge drop in stock prices. On the other hand, as far as the accidents governed by the federal laws are concerned there are restrictions over damages that can be claimed and thus the stock price drops are considerably low.

III. DATA COLLECTION AND PREPARATION

Source of Data :-
www.kaggle.com/saurograndi/airplane-crashes-since-1908

The dataset airplane-crashes-since-1908 data set has 4967 observations and 17 fields with informations on location of crash, Local Time (in 24hour), Flight number, Aircraft type, Total Aboard, Fatalities details and summary that provided a brief description of the accident and cause if known. A review of the dataset suggests that the duration from 07:00 to 21:00 saw a high occurrence of accidents while a large fraction and of accidents related to private air companies (23) and the rest were of different types of air force(11). Based on the comments available in the dataset we have classified the causes into seven categories which are as follows : Weather Error, Air Traffic Control and Navigation Error , Pilot and Staff Error , Unknown , Miscellaneous and Machine Fault. Considering the importance of recency, data from 1981 onwards was analysed

after splitting the entire dataset into three parts of approximately 40 years each and the most recent period was used for analysis.

IV. MODEL TRAINING AND EVALUATION

The renowned IBM too, namely, IBM SPSS Modeler was used to build and train the models. Given the fact that all variables are continuous in nature, following five models were considered for model building. These are: (i) Tree-AS; (ii) XGBoost Linear; (iii) XGBoost Tree;(iv) CHAID; (v)Neural Network. The partitioning used was 75% for training and 25% for testing. A pictorial representation of these is presented below:

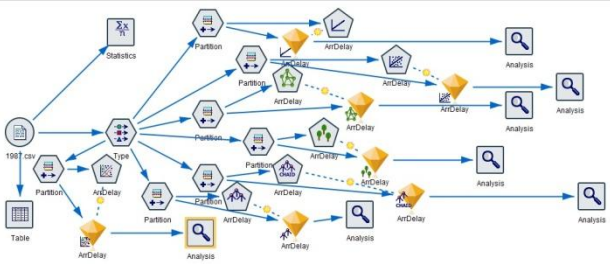


Figure 2: Schematic of Various Models Built

Multiple models were attempted to improve the accuracy.

Neural Network:-The neural network algorithm inspired by the neural network present from the human brain. It consists of a system of neurons which are artificial in nature.Neural networks can attune to changing data, so the best possible result is to redesign the production standard. There will be one or more hidden layer presents between the input and output layers. Numbers of hidden layers depends on the requirement criteria. These layers of neurons compute the weighted input to generate output for the next layer with the help of some activation function and biases.

XGBOOST:-XGBoost means 'Extended Gradient Model', that is used for supervised learning problems. Where we predict a target variable y with the help of training data x. XG Boost follows the decision tree model of machine learning algorithm which uses boosting framework. XG boost model is best for small-to-medium structured or tabular data.

V. RESULTS AND DISCUSSIONS

Use of multiple machine learning algorithms increases the chances of getting better results when no single algorithm is seen to produce optimal results under all conditions. In this specific case, the following four models produce the best results (in descending order): (i) XGBoost; (ii) Neural Network; (iii) CHAID; and (iv) Tree-AS. The result reported relates to the ensemble model that helps address limitations of individual models and also enhances accuracy. These results are presented below as Figure 3:

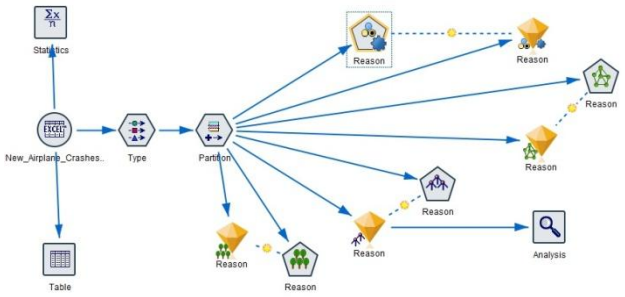
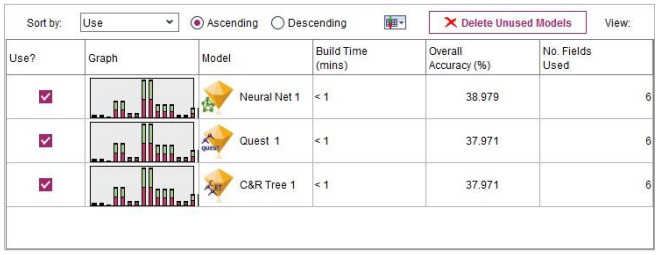
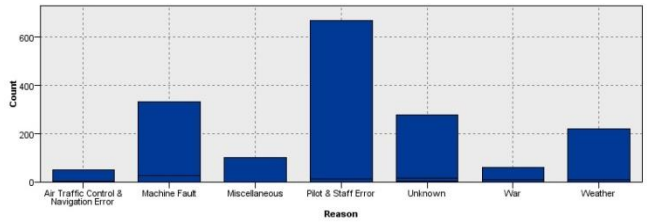


Figure 3: Summary of ModelsUses

Considering the variation in results across different machine learning models, the results of each model in terms of error, standard deviation and correlation are presented below:



. Figure 4: Ensemble Model Results



. Figure 5: Causes Histogram

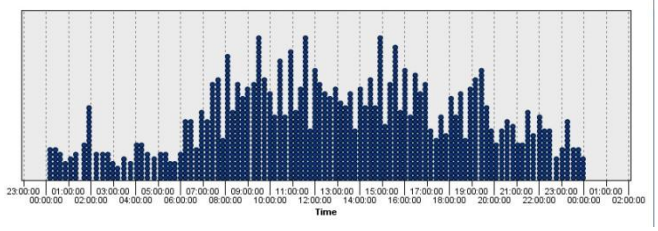


Figure 6: Accident Timing

'Partition'	1_Training		2_Testing	
Correct	848	59.68%	147	39.52%
Wrong	573	40.32%	225	60.48%
Total	1,421		372	

Figure 7: XG Boost Model Results

'Partition'	1_Training		2_Testing	
Correct	528	37.16%	149	40.05%
Wrong	893	62.84%	223	59.95%
Total	1,421		372	

Figure 8: Tree-AS Model Results

Given below is a very brief description of these models listed above:

Figure 11: Predictor Importance

The low importance of time01 could be due to a high correlation between the variables.

VI. LIMITATIONS OF THE STUDY AND SCOPE FOR FURTHER STUDY

The machine learning models used provide an accuracy of over 40% that represents a significant improvement over the random guess of 14% for seven causes. However, there is significant room for improvement through use of added features. This then represents the primary limitation of the study, use of limited features. It also represents an opportunity for further study to enhance the accuracy of prediction.

VII. CONCLUSION

The study suggests that machine learning techniques make it possible to predict the cause of airplane crashes. This could lead to significant savings for planners as well as those investigating air crashes.

REFERENCES

1. Stephens, M.S. and Ukpere, W.I. (2014).An Empirical analysis of the causes of Air crashes. Mediterranean Journal of social sciences,5(2):699,January 2014.Retrived from www.researchgate.net/publication/261613256_An_Empirical_Analysis_of_the_Causes_of_Air_Crashes_from_a_Transport_Management_Perspective
2. Chen Li., V.K. PHUN, T. YAI and M.SUZUKI (2015). The Effect of Air Accidents on Public perception toward an airline.Journal of the Eastern Asia Society for Transportation Studies, V.11:2347, December 2015. Retrieved from www.researchgate.net/publication/294495442_The_Effects_of_Aviation_Accidents_on_Public_Perception_toward_an_Airline
3. T.J. Walker, D.N. Thie., Walker M.G. and K.P. (2014). The Role of Aviation Laws and Legal Liability in Aviation Disasters: A Financial Market Perspective.International Review of Law and Economics,

37:51-65, March 2014. Retrieved from www.researchgate.net/publication/255738443_The_Role_of_Aviation_Laws_and_Legal_Liability_in_Aviation_Disasters_A_Financial_Market_Perspective
4. P.W.Peter (1991). Air Transportation & outlook for the future.From what makes airplanes fly,PP(167-179), January 1991. Retrieved from www.researchgate.net/publication/302190141_Air_Transportation_and_the_Outlook_for_the_Future
5. S.Rao, Shruthi, P. Rao, Sarvesh R. (2018). Airplane Crash Analysis Using LDA.Irjet, v.(05), April 2018. Retrieved from www.irjet.net/archives/V5/i4/IRJET-V5I41086.pdf
6. Y.M. Liao, (2013). Assessments of an Airline cabin safety education program. Evaluation and program planning, 43(C:27-37), October 2013. Retrieved from www.researchgate.net/publication/259002349_An_evaluation_of_an_airline_cabin_safety_education_program_for_elementary_school_children
7. K. Smart (2004). Credible investigation of Aviation accident. Journal of Hazardous material, 111-4, February 2004. Retrieved from www.researchgate.net/publication/8478037_Credible_investigation_of_air_accidents
8. David Young (2014). The crash of American Airlines flight 191. Chicago Tribune, June 2014. Retrieved from www.chicagotribune.com/nation-world/chi-chicagodays-flight191-story-story.html
9. Bureau of Aircraft Accidents Achieved (B3A), on July 12 ,2017, Crashes rate per year, Retrieved from web.archive.org/web/20170728174219/http://www.baaa-acro.com/general-statistics/crashs-rate-per-year/
10. Statistical Summary about Aviation accidents (1959- 2018). Retrieved from www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf
11. Lesson learnt for Aviation Safety, Accident of a Boeing 737 operated by Pegasus Airlines, Turkey, Feb 2020. Retrieved from www.1001crash.com/index-page-description-accident-Pegasus_B737-lg-2-crash-421-pegasus-airlines-boeing-737-turkey-istanbul.html
12. David Gell., A.Troi. and D. Victor (2020). Iran plane crash boeingukraine, The Newyork Time, january 2020. Retrieved from www.nytimes.com/2020/01/07/world/middleeast/iran-plane-crash-boeing-ukraine.html
13. N.S. Alexander (2019). Aviation Metrology at Several Plane Crash, Atmosphere, 10(2):50, January 2019. Retrieved from www.researchgate.net/publication/330704621_Aviation_Meteorology_at_Several_Plane_Crash_Sites

AUTHORS PROFILE

Ved Prakash Gupta is graduating in B.TECH (CSE) from Techno India NJR Institute of Technology, Udaipur. His areas of interest are data analytics, machine learning, programming with python and java. He has good programming skills in C, C++, Python and Java. He has been a team worker, responsible and confident person and a sincere colleague. He is responsible, punctual and creative. He has participated in Digifest5.0 ,software and SiH ,hardware for two consecutive years - Digifest 5.0 software'2018 and SIH hardware'2019. Worked on the Online election system and Traffic controller project as an analyst.

M Sajid Mansoori is graduating in B.TECH (CSE) from Techno India NJR Institute of Technology, Udaipur. His areas of interest are data analytics, NodeJS and competitive programming using Python. He has a good understanding of programming languages such as C, C++, Python and Java. He has been a team worker, responsible and a sincere colleague. He is reliable and innovative.He has worked on Court Case Management System project for SIH 2020 as a NodeJS Developer and also participated in India Innovation Challenge and Design Contest 2018.

Prof (Dr.) Jitendra Shreemali is a graduate from IIT Madras with a post graduate from IIM Bangalore. He is working as Professor of the Department of Computer Science and Engineering of Techno India NJR Institute of Technology Udaipur. He has worked in reputed companies in India & abroad for about a decade and half followed by about two decades of academic/ research/ training experience. He has taught a very wide variety of subjects/courses including operations management, research methodology, and data science besides others. His areas of work include data science, optimization, mathematical modeling and machine learning.

Payal Paliwal received the B.Tech. degree in Electronics and Communication Engineering and M. Tech. in the stream of Digital Communication. She has presented papers in conferences and published more journal papers in the field of digital signal processing and VLSI implementation. She is currently working as an assistant professor in Techno India NJR Institute of Technology, Udaipur.