Improving Aviation Incidents using Association Rule Mining Algorithm and Time Series Analysis

Pierre Pauline R. Abesamis, Remedios de Dios Bulos, and Michelle Ching
De La Salle University, 2401 Taft Avenue, Manila, Philippines

Abstract. Advance technology helps to forecast and show the different hazards associated in the world of transportation systems. Despite the rapid advancement of technology, there are still huge number of aviation accidents and incidents happened. This paper concerns application of various data analytics algorithm to derive a time series forecasting method and frequent accident and incident pattern specifically for Philippine aviation setting for predicting conditions that would increase the likelihood of aviation accidents and incidents involving fatal and non-fatal activities. Data Analytics algorithms are Association Rule Mining using FP-Growth algorithms and Time Series Forecasting methods using Linear Regression, Gaussian Processes, Multilayer Perceptron, and SMOreg are applied to datasets derived from original data obtained from Department of Transportation (DOTr) with its attached agency, Civil Aviation Authority of the Philippines (CAAP) and both incident and accident records stored in “Philippine Aviation Incident Reporting System (PAIRS)” from 2008-2017. The aviation accident data is based on flight information such as aircraft attributes, aviation accident factors, type of occurrences, geographical location, weather conditions and phases of operation. The results lead to prescriptive analytics for development of business rules and help identify pattern that can define aircraft accidents/incidents occurrences. The ability to predict accidents has the main objective of saving lives and will be impactful in cost saving in terms of any aircraft damages. The result created important points in improving the aviation operation and assists in better decision-making to help create policy and future improvement for aviation-safety in the country.

Index Terms— Analytics Value Chain, Aviation Accidents, Association Rule Mining Algorithm, Time Series Analysis

1. Introduction
1.1. Background of the Study
There have been several improvements in safety during flight that led to increase the attention to on-ground risks in the aviation industry that can occur before take-off and after landing. Improvement for the aviation safety has been defined over the years. Most of them are related to technological improvements in aircrafts, avionics and engines that resulted to better aviation safety record [1]. One study according to National Airspace System (NAS) has identified the value of unknown potential hazards and associated risks in improving the system safety. It used data analytics algorithm to create a risk model to identify causal factors [2]. The other is the pilot training. The pilot training has
benefited the improved understanding of human factors and the application of training and regulations. Also, both navigational aids and air traffic management has improved to initiate safer flights with improved weather forecast. In addition, one of the major contributions is improving the safety record that can be traced to thorough investigation within accident and incidents past records which helps prevent for such events to occur. Hence, this reactive approach improves the aviation safety and enhances the data analysis within the various records for both incidents and accidents related to aviation [1].

The aviation data analysis needs to have analysis methods that are effective at finding interesting patterns in the collected data. The analysis of data relies on the past experiences within the aviation accidents and incidents that are caused by multiple factors. Analyzing both accident and incident data can show relationships associated between them. Aviation collection can be from a combination of automated and manual method. However, in this study it will only cater the automated reported events using the PAIR system.

In the Philippines, this is under the Department of Transportation (DOTr) agency with the cooperation of Civil Aviation Authority of the Philippines (CAAP) on which they handle all the aviation accident and incidents. Currently, they have implemented the “Philippine Aviation Incident Reporting System (PAIRS)” program [3]. Its objective is to promote the accident prevention by analysis of safety data. The reporting system facilitates the collection of actual and potential safety deficiencies that include both accidents and incidents report. Also, PAIRS focus on systems, human factors, procedures and equipment and covers all occurrence on which any event can happen. This paper will study the associated factors of frequent incidents in line with the goal of accident prevention. It can help associate rules of frequent aviation accidents to improve the aviation safety and help the aviation community for recommendation for reviewing policy and other plans for improvements.

In this paper, it will use the factors and occurrence for both incident and accident factors based on the report extracted by CAAP in PAIRS system that would be used as data set for this study. It will be tested from the year of 2008 to 2017 that will cover all incident and accident reported. The data used for this research study were collected from Open Data Philippines. Open Data Philippines is an online platform that allows government agencies to be transparent with their actual data and be visible to its citizens. Despite seeing the results of the collected data, DOTr still have unconsolidated data and needs data integration with its different attached agencies such as in CAAP to analyze the various data sets. Thus, today there is a need to learn from past data to apply prediction of future occurrences of events using the past data. Overall, this study can help them decide for the process improvement and policy enhancement for the enhancement of aviation safety.

1.2. Objectives of the Study
This research study aims to help CAAP and DOTr in consolidating and analyzing data in the past years for both incident and accident aviation events by having quantitative data analytics. The main objective of this paper is to identify associated factors of frequent incidents in line with the goal of accident prevention. Furthermore, it will identify significant accident factors and evaluate the trends of aircraft accident and incident events reported in the years of 2008 to 2017. This information is helpful for the agency to lessen frequency of accidents and to maximize the contributing variables in promoting aviation accident safety.

It will be done using Association Rule Mining using the FP-Growth algorithm in RapidMiner data science tool to formulate associating rules of pattern for aviation incidents that covers both incidents and accidents. Moreover, time series forecasting using Waikato Environment for Knowledge Analysis (Weka) machine learning workbench will be used to conduct a one-year predicted data and result with be compared to determine the most accurate prediction using Linear Regression, Gaussian Processes, Multilayer Perceptron, and SMO Regression. This is useful to understand its pattern to extract valuable information that can help recommend aviation-safety strategies.

1.3. Scope and Limitation of the Study
Open Data Philippines has different kinds of aviation data sets provided by CAAP as attached agency of DOTr that handles all the incidents for both aviation incidents and accidents. The focus of this study will only for reported aircraft incidents and accidents using the Philippine Aviation Incident Reporting System (PAIR) program within the years of 2008-2017. Initially, the data sets will only cover the standard aircraft incident and accident reports and will not be covering the data that were not reported and not retrieved in the system. Also, it will cover all location within the Philippines both in cities and provinces and all of the incident and accident factors would be considered in the study.

Moreover, the researcher chose RapidMiner as the data science tool to be used for conducting the Association Rule Mining using the Frequent Pattern (FP-Growth) Algorithm to find association rules of frequent aviation accident pattern or items sets mining in market basket analysis and to be able to make variations [4]. Another tool is Weka’s Forecast package will be used for time series forecasting using Linear Regression, Gaussian Processes, Multilayer Perceptron, and SMO Regression to determine the most accurate one-year forecast for the identified PAIRS factors (human, equipment, procedures and systems). In addition, it will show the descriptive part using the MS Excel 2016’s Analysis ToolPak that allows data analysis tools to have statistical and engineering analysis in data visualizations and descriptive analysis. It is used to provide the visual interface on a workflow and user-friendly environment. It will also be integrated within R programming language that shows a complete visual workbench in delivery appropriate result and report. Aside from the above tools, MS Excel 2016’s Analysis ToolPak together with R programming language. It allows data analysis tools to have statistical and engineering analysis in data visualizations and descriptive analytics.

2. Review of Related Literature Work
2.1. Related Research Studies
Data Analysis in accident is very helpful since it can reveal the relationship between the different types of attributes that can lead to an accident. Several types of road accidents such as in road, traffic and airplane have different ways of comparison and nature since they are all unpredictable. However, this study will only focus on aviation accidents.

Data mining is a process on which it will help create useful and understandable knowledge within the data [5]. Thus, it should be used to examine the aviation accident data set and particularly it will focus on Association Rule Mining and Time Series Forecasting [6]. Based on literature, Aviation safety data analysis commonly face issues of finding analysis methods that are effective at finding interesting patterns within the collected data [7].

2.2 Association Rule Mining
Association Rules Mining is one of the most common and very important aspect of data mining for it is being used to find some interesting associations or patterns within correlation relationships between item sets within a huge amount of data [8]. It is commonly used to identify frequent item sets and is one of the most powerful tools for technology and a useful step in the application of association rules mining [9]. Also, it is a technique on which it patterns to the market basket analysis of market data and focus is on identifying the goods on which the customer prefers to buy together more frequently. Therefore, it produces several sets of rules that can improve the correlation among different set of attributes in data set. It has interesting measures like the support and confidence that results to have stronger rules. The support value shows the frequency of occurrence of a rule on a data set and the confidence illustrates about the reliability of a rule. Thus, a high confidence and support value is the main interest on any association rules. In addition, clusters analyze using association rule mining find the correlation with data attributes. And each cluster is analyze using the association rule mining algorithm like FP-Growth (frequent-pattern growth) algorithm [10].

FP-Growth algorithm is the improved version of the Apriori algorithm that has been started by Jiawei Han. It compresses the data sets to a FP-tree (frequent-pattern tree) and scans the database twice but still maintains the information of associations between item sets. The compressed database is divided into sets of condition database and each condition associates with a frequent item. Also, it does not produce a candidate item sets in mining process, but it greatly improves its mining efficiency.
However, the FP-Growth algorithm needs to have a FP-tree that contains all data sets which occupies higher requirement in memory space [5].

2.3. Time Series Forecasting

Time series method is being used to analyze any statistical characteristics of aviation accident data. In most statistical data, there are logical events that can show the periodicity and timing that affects the statistics by variety of factors. Time series is based from past historical data that is shown in a regular basis that is combined together with influencing factors of data to help make future predictions of data. In this paper, it is an important feature that contributes to the significance of aviation accidents trend analysis [11].

Time series forecasting analysis composed of description of the component’s movements of the present data. There is a graphical representation for example: time plot reveals quantitatively the presence of long-term trend, cyclical, seasonal and irregular variation. Time series analysis also investigates the following factors and in decomposition of time series into basic component movements. The time plot enables to construct the long trend term curve that can be used as (a) method of freehand, (b) least squares methods, (c) moving average method and (d) semi-average method [12].

For this study, the research was examined and applied in various forecasting techniques used in Weka. Weka is a tool to compare various prediction functions and found that SMOreg (SMO regression) function showed the ability for the most realistic and accurate aviation accident prediction other than the existing methods such as the Linear Regression, Multilayer Perceptron and Gaussian processes. Some of the researchers also recommended to use either SMOreg or Gaussian processes to determine the most accurate prediction.

3. Methodology

In this section, it explained how the collection and preparation of aviation accident dataset were done. It provides a discussion on how the data were analyzed and used for statistical analysis. Also, it includes how data mining models were created to answer the objectives of this research. This study will use various tools namely: (a) RapidMiner as data analytics tool to demonstrate Association Rules Mining particularly the FP-growth algorithm, (b) Weka’s Forecasting package for the Time Series forecasting using Linear Regression, Gaussian Processes, Multilayer Perceptron, and SMOreg, and (c) MS Excel 2016’s Analysis ToolPak to show the descriptive analytics, and visualizations. The below section discusses the study’s operational framework and algorithm used for the whole study.

3.1. Operational Framework

The Analytics Value Chain as shown in Figure 1 is the operational framework used for this research study.

![Figure 1. Analytics value chain.](image)

In the Analytics Value Chain, it begins with the curate phase on which it covers the process of collection, storage, management and cleaning of data. The second phase is the summarize for it includes the charts, tables and visualization techniques to summarize the clean version data sets. In the descriptive part, it is used to know the past and present trends while the predictive focuses more on the forecasting outcomes of possible events. Those data can be helpful upon creating possible business rules in the future. The prescriptive shows recommendation of possible actions to do based on all the data gathered as discussed by IBM Smarter Analytics (2013).

3.2 Data Collection

To test and evaluate the aviation accident prediction and forecast models, the collected historical accident and incident data were gathered from the Civil Aviation Authority of the Philippines (CAAP)
on which it is an attached agency of Department of Transportation (DOTr) that handles all the aircraft issues and aviation accidents. With the help of the “Philippine Aviation Incident Reporting System (PAIRS)”, they can extract an excel reports that can be useful collection of information on actual or potential safety deficiencies that is being captured by the mandatory incident reporting system [3].

The aviation accident dataset consists of all the location in the Philippines and within the focus area of the PAIRS which are systems, human factors, procedures and equipment. It also supports all the areas covered within the PAIRS program which include departure/in route/approach/landing operations, aircraft ground operations, movement on the airport, fueling operations, safety-related passenger, airport conditions or services or any aviation-safety related issues that were not categorized. The duration of years is from 2008-2017. Initially, a total of 1,214 data were used to create the model. However, all the data set that were used in the study were only gathered in the Open Data portal on which it shows the available government data open for public used.

3.3 Data Preprocessing

Data pre-processing is the important step in data mining process. It has the phrase of “Garbage In, Garbage Out” concept that is available to data mining and machine learning. Analyzing data that is not cleaned thoroughly produces a lot of problems that can lead to unreliable results. With this, quality of data is being done first before going to an actual analysis of data. Data preparation purpose is filtering steps that can take huge amount of time. The data preprocessing has data cleaning, data integration, data transformation and data reduction. The results of this can now be used for data mining [13].

3.3.1 Data Cleaning

Data would be analyzed using the data mining techniques and at first it can be incomplete such as lack of attributes values, attributes of interest, noisy (consists of errors) and inconsistent (discrepancies of other data) [13]. In this section, there were some missing and inconsistent values within the attributes of factors, weather condition, phase of operation, impact and type of occurrence of the accident. Instances that lacked those attributes were filled in from the dataset. In addition, the duplicate instances were solved. The date was separated into (month, day and year) so that the year can be used as separate attribute. The place of occurrence was also classified into a province attribute so that some of the redundant and duplicate entries can be grouped into one location. Moreover, some of the type of occurrence have new occurrence that is not relevant to the normal group which makes it a noisy data. Also, report type, status and report attribute were removed since they are not going to be helpful in the model.

3.3.2 Data Integration

Data integration involves combining of data from multiples sources. In this case, there are separate reports created by the PAIRS system. Reports produced were categorized into accidents and incident and year per year generated reports. In this paper it integrates the separate excel sheets that has several reports within accident and incident category and extracted per year basis.

3.3.3 Data Transformation

The data transformation, the data are transformed into consolidated forms needed for data mining. Within the data set, there is the generalization of the data on which the low level or “primitive” (raw) data replaced by higher level concepts through the use of concept hierarchies [13]. In this case, the categorical attributes like the place of occurrence and the type of occurrence where generalized to higher level concept which are “province” and “group of occurrences.” Hence, it will help within the analyzing data and producing better patterns for aviation accident prevention.
Table 1. Key aviation accident attributes table.

| Attribute Name          | Description                                      | Data Type   | Value                                                                 |
|-------------------------|--------------------------------------------------|-------------|----------------------------------------------------------------------|
| factors                 | PAIRS program key indicator                      | Nominal     | Systems, human factors, procedures and equipment                     |
| group_of_occurrence     | The type of occurrence that aircraft accident/incident occur | Nominal     | Ditch, Fuel Starvation, Engine Failure, Excursion, Crash, Overshoot, Undershoot, Tail Strike, Gear Failure, UFIT, Stall, Collision, Collapse, Fire, Propeller Strike, Loss Control CFIT, Nature, Abort, Tire Failure, Taxiway Excursion and Others |
| phase_of_operation      | Identifies what operation phase the accident happened | Nominal     | Landing, Inflight and Take-Off                                      |
| weather_condition       | Identifies the weather condition the time the accident occurred | Nominal     | Visibility, Wind and Cloud                                          |
| impact                  | Determines the impact damage of the accident      | Nominal     | Fatal and non-fatal                                                 |
| province_of_occurrence  | Shows the location of the accident                | Nominal     | Cebu, Baclaran, Zambales, Manila, Bulacan, Palawan, Bohol, Davao, Aklan, Pampanga, Iloilo, Las Pinas, Leyte, Quezon City, South Cotabato, Tarlac, Albay, Paranaque, Subic, Baguio City, Pangasinan, Masbate, Antique, Bukidnon, Surigao Del Sur, Kalinga, Pasay City, Samar, Nueva Vizcaya, Mindoro Occidental, Tacloban, Ilocos Sur, Aguasan Del Norte, Cavite, Romblon, Tuguegarao, La Union, Bicolod, Dumaguete and Maguindanao |
| year                    | The year the accident happened                    | Nominal     | 2008-2017                                                             |

3.3.4 Data Reduction
The concept of data reduction in this study applied the reduction of the dimensions (total number of attributes) [13]. It used the dimension reduction on which it removed the irrelevant, weakly relevant and the redundant attributes were detected and removed. It removed the attributes status and reporting for it would not be useful in developing the model. Also, it now depends on the discretization and concept hierarchy generation on which the raw data values for attributes are replaced by ranges or higher conceptual levels. These are both done in the type of occurrence and place of occurrence attribute within the dataset. The result of data reduction is presented in Table 1.

3.3.5 Data Mining
Data mining is referred to as knowledge discovery in databases (KDD). It is the process for sorting within huge datasets to identify patterns and establish possible relationships to solve a particular problem in doing data analysis. Also, data mining is being used to predict future trends. The data mining tools is being used to perform data analysis that uncovers important data patterns and
contributes to help in business strategies [14]. The data mining technique used in the paper is Find Similar and is designed to search a collection of aircrafts incidents and accidents to find the most similar to selected accidents. It would be useful in determining if similar incidents or accidents occurred before and if how they are addressed. In the bottom section of the study, it will apply the Association Rules Mining and Time Series forecasting.

4. Data Analytics
Data Analytics is currently being used to describe the analysis of large volume of data on which it shows unique computation tool in handling big amount of data in any organization [15]. It can be both qualitative and quantitative data analysis in ensuring productivity that is achieved using extracting, categorizing and analyzing data in finding meaningful patterns that can help the organization. With this, the researcher aims to study the data set for aviation accident under the PAIRS program in the Philippines. It will apply data analytics and will follow the Analytics Value Chain Operational Framework and will be discussed within the paper [16].

4.1 Data Curation and Summary
The focus of the research study is on the aviation incident and accident data sets reported in the year of 2008-2017 that extracted from the “Philippine Aviation Incident Reporting System (PAIRS)” program. It includes the attributes Factors, type of occurrence, phase of operation, weather condition, impact, year and province of occurrence. Nevertheless, there are duplications with the type of occurrence and province of occurrence which are cleaned and group together to have a standard name for each. The cleaning of data is done manually using MS Excel and resulted to having Table 2 for PAIRS factors and Table 3 for types of occurrence.

Table 2. PAIRS factor Data Set Summary

| PAIRS factors | Fatal Accidents | Non-Fatal Accidents |
|---------------|----------------|---------------------|
| Equipment     | 144            | 210                 |
| Human factors | 120            | 168                 |
| Procedures    | 96             | 174                 |
| Systems       | 156            | 150                 |

Table 3. Types of Occurrence Data Set Summary

| Type of Occurrence | Fatal Accidents | Non-Fatal Accidents |
|--------------------|----------------|---------------------|
| Abort              | 0              | 6                   |
| CFIT               | 0              | 6                   |
| Collapse           | 36             | 12                  |
| Collision          | 36             | 24                  |
| Crash              | 12             | 54                  |
| Ditch              | 12             | 6                   |
| Engine Failure     | 204            | 174                 |
| Excursion          | 48             | 84                  |
| Fire               | 6              | 30                  |
| Fuel Starvation    | 0              | 18                  |
| Gear Failure       | 60             | 120                 |
| Loss Control       | 6              | 6                   |
| Nature             | 6              | 0                   |
| Other              | 0              | 42                  |
| Overshoot          | 12             | 18                  |
| Propeller Strike   | 12             | 12                  |
4.2 Descriptive Analytics

The Descriptive analytics uses the business intelligence and data mining that answers the question “What has happened?” it mines data to provide the trending information on past and current events and can give a real estate of what happened that is in need for future movements. Using the descriptive data accumulated over time, predictive analytics utilizes models over time. To summarize the important data that is related to the aviation accidents and incidents, below are the different tables, figures and visualizations [17].

The collected data shows in Figure 2 presents the total number of aviation accident and incidents within the span of the year 2008 to 2017. It shows that there are big number of non-fatal impacts that total of 58% against the fatal impact of 42%. Within the visualize histogram, there are a lot of accidents that happened in 2013 that cause non-fatal accident while on the year 2011 it shows higher value fatal aviation accidents.

| Event             | Non-fatal | Fatal |
|-------------------|-----------|-------|
| Stall             | 24        | 12    |
| Tail Strike       | 6         | 24    |
| Taxiway Excursion | 6         | 12    |
| Tire Failure      | 12        | 12    |
| UFIT              | 18        | 6     |
| Undershoot        | 0         | 24    |

Figure 2. Total aviation accidents impacts.

Another important attribute used in the data is the PAIRS factors. In the below histogram, it shows the comparison of the different PAIRS factors which are the equipment, human factors, procedures and systems that results to both fatal and non-fatal impacts. The diagram in Figure 3 shows that the top PAIRS factor is the equipment that leads to non-fatal impact.

Figure 3. PAIRS factors (per impact).
In Figure 4, it displays the top five aviation accidents which are the *Collision, Crash, Engine Failure, Excursion and Gear Failure* that happened on all phases of operation covering *Inflight, Landing and Take-Off*. It reports that the “Engine Failure” is the top cause of aviation accidents that normally happens during the “Landing” phase of operation in any reported aircraft accidents.

![Figure 4. Top five aviation accidents occurrence (per phase operation).](image)

For figure 5, it displays the Aviation Accidents Phase in the whole year from 2008-2017 and covers all phases of operation on which the aviation accident occurred. The diagram shows that the year of 2013 during the Landing phase of operation is the most frequent number of aviation accidents. Compared to the three phases of operation, *Take-Off* is the least frequent phase of operation in all the years.

![Figure 5. Aviation accidents phase (per year).](image)

For figure 6, it shows the Aviation Accident Weather Conditions can be classified into cloud, visibility and wind. In the pie chart, it shows that the frequent number of aviation accident happens in the windy condition that has 37% and the least frequent accident happens during a cloudy weather which only consists of 30%.

![Figure 6. Aviation accident weather conditions.](image)
In Figure 7 below shows the top five locations where the aviation accidents occurred in between the years of 2008-2017. The top five places are: Aklan, Bulacan, Davao, Manila and Palawan. The dataset includes all the fatal and non-fatal accidents, all the pairs factors, and all-weather conditions. The results show that among the five, the peak of aviation accident happened in the place of Davao and it happened in the year of 2011. It concludes that Davao has consistent increase with the number of accidents every year.

4.3 Predictive Analytics

The Predictive Analytics uses statistical models and forecasts to answer the question “What could happen?” It gives answers that move beyond using historical data as the primary basis for all decisions rather it helps the leaders to anticipate likely scenarios that plan ahead than just reacting to a previous event that occurred. It utilizes models for predicting events. Nevertheless, it does not recommend any actions. This study would be using forecasting and simulation to help enhance the views of the decision makers to make more informed decisions [17]. It will apply the concepts of Association Rule Mining and Time Series Forecasting that would clearly define below.

4.3.1 Association Rules Mining

In this paper, the frequent pattern growth approach (FP growth) is used on which it finds association rules (frequent pattern) and modify it to determine the frequent item sets [4]. The objective of this is to find the patterns in the given aviation accident and which of these association rules are form based on the its minimum support, confidence, and lift. It then helps address the objective of the paper which is to identify associated factors of frequent incidents to the goal of accident prevention. These may assists the value of the reports produced by PAIRS as their incident reporting system and can help contribute insights that can be part of the State Safety Program (SSP) and Safety Management System (SMS) of every aviation organization certified by CAAP [3].

The attributes used in the association rule mining are the factors, year, type_of_occurance, phase_of_operation, weather_condition, impact and province_of_occurance. All the instances are in a nominal type. To be able to subject the instances to association rule mining, the values required is binomial. Within the preprocessing these data, all the nominal values are converted into numerical version then binomial using the operator commend that is available in the datamining tool, RapidMiner.

The RapidMiner was the tool used for finding the frequent patterns. In validation of the patterns created, there is the minimum support, confidence and lifts measures that was used. The support describes how frequent these patterns occur in the dataset. For the confidence, it illustrates about how strong those associations are while the lift, is used to test the interestingness in determining if the values of the antecedent have a positive effect on the consequent. Thus, the lift value that is greater than 1 will result to a more positive correlation.
Table 4. FP-Growth results for aviation accidents frequent patterns.

### Conclusion: Fatal (impact)

| Frequent Patterns                                | min support | Confidence | lift  |
|-------------------------------------------------|-------------|------------|-------|
| Phase_of_operation = Landing, type_of_occurrence = Engine failure → Fatal | 0.113       | 0.548      | 1.293 |

### Conclusion: Non-Fatal (impact)

| Frequent Patterns                                | min support | Confidence | lift  |
|-------------------------------------------------|-------------|------------|-------|
| Phase_of_operation = Landing, weather_condition = Visibility → Non-Fatal | 0.128       | 0.591      | 1.025 |
| Phase_of_operation = Landing, factors = human factors → Non-Fatal | 0.084       | 0.607      | 1.053 |
| Phase_of_operation = Landing, weather_condition = Cloud factors → Non-Fatal | 0.118       | 0.632      | 1.096 |
| Phase_of_operation = Landing, type_of_occurrence = Gear Failure → Non-Fatal | 0.084       | 1.180      | 1.096 |
| Phase_of_operation = Landing, factors = procedures Failure → Non-Fatal | 0.103       | 0.700      | 1.215 |
| Factors = equipment, province_of_occurrence = Davao → Non-Fatal | 0.069       | 0.700      | 1.215 |
| Weather_condition = Visibility, factors = equipment → Non-Fatal | 0.074       | 0.714      | 1.239 |

In this approach, Table 4 shows the FP-Growth was used to discover the best association rules with interesting aviation accidents attributes. The threshold for min support is set to 0.15 and confidence level is set to 0.5 accordingly. This results to have 8 frequent patterns with the impact as the consequent. As seen in the table, there is only one frequent pattern for fatal impact this can imply that this type of impact is not normally happening. It can be seen that non-fatal incidents produce 7 frequent patterns that associate in different attributes from the dataset. Most of the patterns shows that the aviation accident often occurs during the Landing phase of aviation operation and it is matched with different PAIRS factors. Hence, for the non-fatal impacts the greater positive correlation based on lift has the value of 1.239 resulted from the combination of Visibility and equipment as factors that leads to non-fatal accidents.
4.3.2 Time Series Forecasting

For this paper, time series forecasting was used with the help of Weka datamining tool. The results were extracted and compared to different algorithm namely: Linear Regression (LR), Gaussian Processes (GP), MultiLayer Perceptron (MP), and Sequential Minimum Optimization Regression (SMOreg) to predict the future trends for the aviation accident for one year in the Philippines. To measure its accuracy, it used the concept of Mean Absolute Error (MAE). MAE is a forecasting accuracy measure to help choose the best-suited method for the aviation accident dataset extracted. The methods used only to look for one year forecast.

The experimental results shows that one year periodicity generated the minimum MAE and it shows the percentage error in each method of the forecasted year which is seen in table 6. By comparing the results, the SMOReg offers the ability to predict the number of aviation accidents in most PAIR factors more accurately compared to other techniques (see table 5). The PAIR factors that got the least MEA are systems and procedures. The forecast value using the SMOreg method is seen in the table 6. The predicted aviation accidents range is 8-30 and the computed MEA for prediction is 0.88 which is low. Hence, it implies that the forecasting method generated better performance. It can be used to reduce the error percentage in predicting the future aviation accidents and it increases the chances of the aircraft officers to predict more accurately.

| PAIR factors     | LR   | GP   | MP   | SMOReg |
|------------------|------|------|------|--------|
| Equipment        | 30.19| 25.65| 24.59| 27.61  |
| Human factors    | 6.44 | 17.11| 13.07| 12.09  |
| Procedures       | 24.6 | 24.74| 22.87| 31.33  |
| Systems          | 7.24 | 13.5 | 30.79| 8.16   |

Table 5. Predicted PAIR factors values (one year).

| Percentage of Error  | LR     | GP     | MP     | SMOReg |
|----------------------|--------|--------|--------|--------|
| Correlation Coefficient | 1      | 0.83   | 0.98   | 0.99   |
| Mean absolute error  | 3      | 17.5   | 4.06   | 0.88   |
| Root Mean squared error | 5      | 21.68  | 5.09   | 1.55   |
| Relative absolute error | 0      | 55.84  | 12.95  | 2.8    |
| Root relative squared error | 0      | 57.01  | 13.4   | 4.08   |

Table 6. Percentage of error.

4.4 Prescriptive Analytics

The Prescriptive Analytics uses optimization and simulation to answer the question “What should we do?” It is the next method based on the Analytics Value Chain Operational Framework that suggests actions based on the descriptive and predictive analysis of complex data. It results to development of business rules, organization models, comparisons and optimizations based on the outcome of the Predictive Analytics. This study will help prevent occurrence of aviation accidents in the country [17].

In Association Rule Mining, the frequent pattern growth approach (FP growth) algorithm finds the common occurring aviation accident patterns. The results showed that non-fatal impact aviation accidents has a high probability to be from equipment and procedures failures that is associated with frequent engine failures (Landing – Gear Failure – Equipment). The implications of the high rate of failures during the landing phase of operation together with high rate of failures in equipment factor
(engine and gear failure) to avoid aviation accidents with fatal impact indicate that solving this problem alone could remedy a majority of aviation accident both for non-fatal and fatal accident impacts.

Another result showed that aviation accidents with non-fatal impacts are more likely to happen in the future during the aircrafts landing phase of operation and the weather condition remains visible and not raining. This is the second most frequent occurring pattern of aviation accident in the dataset is in landing phase with weather circumstances (Landing- Weather - Human Factors). This implies that the impact of accident was not fatal for a visible weather and human factors is the only element that can lead to aviation accident. The importance of understanding the implications of weather and formulating educated in-flight decisions regarding weather must be understood in different levels. In several instances, the pilots without training found themselves in circumstances of deteriorating weather. Instead of choosing alternate pilot, pilots flew into the weather, often lost situational awareness, became spatially disoriented or attempted to land while on unstable approach. In certain circumstances, attempting to land from unstable approach could be considered intentional non-compliance. However, nearly all of the weather-related cases, poor decision-making was demonstrated [18].

For the application of time series forecasting, the SMO regression was used to forecast the number of aviation accidents in all PAIRS factors more accurately compared to the other techniques such as Gaussian processes, linear regression, and Multilayer Perceptron. The outcome of these facilitates the prevention and mitigation of aviation accidents in the future time. Thus, with the help of this study it can help decrease the number of aviation accidents in the Philippines and increase the effectiveness and efficiency of aviation accident data analysis.

5. Conclusion
Aviation community in the Philippines still faces challenge on how to maximize the data produced by PAIRS both for incidents and accidents to help prevent aviation accidents. With the help of Open Data, it makes the aviation incidents and accidents scenarios be opened to the public for different span of years. This research study proves that there are associated factors of frequent incidents to the goal of accident prevention. It is important to consider all the associated factors because it has the vital information on the cause of aviation accidents and to help provide the impacts of the said accidents. It uses the Association Rule mining with the help of frequent pattern (FP-growth) algorithm to find the frequent aviation accidents. For the time series forecasting methods, the SMO regression was found to forecast the number of accidents on each PAIR factors accurately. The result created important points in improving the aviation operation and assists in better decision-making to help create policy and future improvement for aviation-safety in the country.

Acknowledgments
Author P.P.R. Abesamis would like to express her gratitude to Dr. Remedios de Dios Bulos, co-author and professor under the course Information Systems Planning, Design, Analysis and Databases, for the opportunity and data analytics knowledge that has been useful in the research paper; to Mr. JP Acuna for recommending the challenge areas that government faces in handling big data; to co-author Ms. Michelle Ching for guiding me with Data Analytics and research principles; to Mr. Neil Datu and Ms. Michelle Cortez, for sharing their knowledge and insights on Data Analytics together with the tools of Weka and RapidMiner and motivating me to finish it; Thank you De La Salle University Manila that provided financial assistance to present this in a conference and publish it in a journal and last but not the least, to our Almighty Father for providing us strength, wisdom, knowledge and motivation to finish this meaningful research study.

References
[1] J. S. S. C. K. Z. Clinton V. Oster, "Analyzing aviation safety: Problems, challenges, opportunities," Research in Transportation Economics ELSEVIER, pp. 148-164, 2013.
[2] J. T. Luxhoj, "Predictive Safety Analytics for Complex Aerospace Systems," Procedia Computer Science (Elsevier), pp. 331-336, 2013.
[3] CAAP, "Republic of the Philippines Department of Transportation (CIVIL AVIATION AUTHORITY OF THE PHILIPPINES)," Civil Aviation Authority of the Philippines, [Online]. Available: http://www.caap.gov.ph/?page_id=2180. [Accessed 10 December 2018].

[4] P. N. G. P. A. S. Rakesh Kumar Soni, "An FP-Growth Approach to Mining Association Rules," International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 2, no. 2, pp. 1-5, 2013.

[5] S. Y. J. L. a. M. Z. Yi Zeng, "Research of Improved FP-Growth Algorithm in Association Rules Mining," Scientific Programming, vol. 2015, 2015.

[6] P. Tiwari, Accident Analysis by Using Data Mining Techniques, 2017.

[7] E. B. a. P. O. Zohreh Nazeri, "Experiences in Mining Aviation Safety Data," in ACM SIGMOD, 2001.

[8] D. Fengyi and L. Zhenyu, "An ameliorating FP-growth algorithm based on patterns-matrix," Journal of Xiamen University (Natural Science), vol. 44, pp. 629-633, 2005.

[9] Y. Yang and Y. Luo, "Improved algorithm based on FP-Growth," Computer Engineering and Design, pp. 1506-1509, 2010.

[10] S. Kumar, "Analyzing Road Accident Data Using Association Rule Mining," IEEE, 2015.

[11] D. Z. a. K. Jianga, "Application of Data Mining Techniques in the Analysis of Fire Incidents," International Symposium on Safety Science and Engineering in China, 2012.

[12] Balogun.O.S., "International Journal of Advanced Research," USE OF TIME SERIES ANALYSIS OF ROAD ACCIDENT DATA IN LAGOS STATE, vol. 2, no. 5, pp. 1045-1059, 2014.

[13] D. A. Arora, "Data Mining Techniques and Tools for Knowledge Discovery in Agricultural Datasets," IASRI, 2011.

[14] J. H. a. M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2000.

[15] Informatica, "What is Data Analytics?" Informatica, 2018. [Online]. Available: https://www.informatica.com/services-and-training/glossary-of-terms/data-analytics-definition.html#fbid=hYXDmWmJvom. [Accessed 12 December 2018].

[16] "Data Analytics," technopedia, 2018. [Online]. Available: https://www.techopedia.com/definition/26418/data-analytics. [Accessed 30 November 2018].

[17] IBM, "IBM Software Thought Leadership White Paper," Descriptive, predictive, prescriptive: Transforming asset and facilities management with analytics, 2013.

[18] B. N. Branham, "Analysis of Fatal General Aviation Accidents Occurring from Loss of Control on Approach and Landing," EMBRYRIDDLE Aeronautical University Scholarly Commons, August 2013.