Application of Predictive Modelling to Determine Factors Influencing Flight Delays at Jomo Kenyatta International Airport, Kenya

Gachoki PK[1], Munyiri LM[2] and Moses Mburu[3]

[1]*Department of Physical Sciences, Chuka University, P.O Box 109-60400, Chuka, Kenya*
[2]*Department of Pure and Applied Sciences, Kirinyaga University, P.O Box 143-10300, Kerugoya, Kenya*
[3]*Kenya Medical Research Institute, PO Box 230-80108,Kilifi, Kenya*

Abstract: - **To improve decision making accuracy in any given airport requires selecting important variables that can be used in building a predictive model. The choice of appropriate independent variables will improve the model precision. However, the choice of the independent variables depends on the data that is recorded by the airlines and airports on the flights. The airport managers would want to understand the key factors behind flight delays. It is thus important to make a comparison on the mostly used factors in many modelling studies of flight delays and the factors that influence flight delays at Jomo Kenyatta International Airport (JKIA). The factors mostly used are the weather and the flight features. The factors available at JKIA include; the day of the flight (that is, Monday to Sunday), the month (that is, January to December), the airline, the flight class (that is, domestic or international), season (that is, summer (March to October) or winter (October to March), capacity of the aircraft, flight ID (tail number) and whether the flight had flown at night or during the day. The data used was obtained from Kenya Airports Authority for the JKIA flights for the period from March 2017 to March 2018. The analysis was done using R-Gui statistical software. Descriptive statistics were generated to give a general overview of how the above factors influenced flight delays at JKIA. Logistic model was then fitted to demonstrate how the factors could be applied in predicting flight delays. This model was also used to extract the significant factors in predicting flight delays. The selected factors were also compared on the performance they yielded in modelling as compared to features which had been used in other studies. The results revealed that the significant factors were days of the week, months, flight class and capacity. Modelling using these factors yielded models with average F1 score of 76.95%. This was better performance when compared to results from another study that used predictive features such as: the previous aircraft arriving late, weather, and departure timeand achieved an average F1 score of 58.7%. Another study predicted airline delays using flight departure times, and weather conditions. Their prediction algorithms achieved F1 score of 56.6%. This shows that the factors that influence flight delays at JKIA improves the performance of predictive models of flight delays.**

Key Words: **Flight Delays, Independent Variables, Predictive Modelling, Factors that Influence Delays, F1 score.**

I. INTRODUCTION

In developing prediction models for flight delays, the choice of the independent variables depends on the data that is recorded by the airlines and the airports. Factors such as distance covered by an aircraft, the day of the flight and the scheduled departure time were considered when developing a multiple regression model to predict flight delays (Burgauer & Peters, 2000). In that study, those factors were found key in predicting flight delays. Time duration and number of airports were used in developing probability models for flight delays (Mueller & Chatterji, 2002). The probability models developed using these factors were the normal distribution and the poisson distribution. An econometric model that was used to study flight delays used various variables such asairline and airport of destination, frequency, aircraft size, occupancy rate and fare(Hsiao & Hansen, 2006). This model was used to make a study on flight cancellation. A correlation analysis has been done between flight delays and capacities of airports (Ison*et al*., 2015). The results from this analysis indicated that a correlation exists between flight delays and airport capacities. However, these results called for further investigation since it was not shown with certainty that one variable caused another one to happen. There was a possibility of another variable that could have caused the correlation. Factors such as the airport capacity, departure delays and arrival delays have been used to model the cost of a delayed flight using advanced analytical techniques in operation research such as simulations (Zou & Hansen, 2012). In a study by Kalliguddi & Leboulluec (2017), predictive modelling of aircraft delay was done using variables such as, national air system delay, departure delay, carrier delay, taxi in and taxi out, weather delay, late aircraft delay, distance and security delay. Lawson & Castillo, (2012) used a dataset of flights and included several years, which resulted in an impressive number of data points. They limited their features to weather data only (33 feat ures only in the end) obtaining 40% recall only. A study by Sridhar et al. (2009)focused on weather related delays but used more sophisticated features such as key airspace metrics and also took into account the delay of previous flights to predict a new delay. Predicting flight on-time performance was done using predictive features such as: the previous aircraft arriving late, weather, and departure time and achieved an average F1 score of 58.7%. (Arjoun*et al*., 2013). Bandyopadhyay & Guerrero (2012) predicted airline delays usingflight departure times, and weather conditions. Their prediction algorithms achieved F1 score of 56.6%. This clearly shows that for different studies, different factors will be used for predictive modelling. This will mainly be influenced by the data collected from the airlines and airports. This data is also dependent on the mechanisms of data collection put in place by the airlines and the airports. The factors that influence flight delays at JKIA have not been determined. The objective of this study is to determine the factors that can be used in developing prediction models for flight delays at JKIA. Further, a comparison is made on how these factors affects the performance of the predictive models of flight delays.

II. METHODOLOGY

The data used was collected from Kenya airports Authority for the JKIA flights for the period from March 2017 to March 2018.The recorded variables included; the day of the flight (that is, Monday to Sunday), the month (that is, January to December), the airline, the flight class (that is, domestic or international), season (that is, summer (March to October) or winter (October to March), capacity of the aircraft, flight ID (tail number) and whether the flight had flown at night or during the day. The entire data set had 20000 flights after data cleaning, filtering and imputations. A flight delay was calculated by getting the time difference between the scheduled time and the actual movements time of specific aircrafts. However, this difference was classified as a flight delay if and only if it was 15 minutes or more. This is as per the international standards as set out by International Civil Aviation Organization (ICAO). The analysis of the factors that influence flight delays at JKIA was done by generating descriptive statistics in form of tables and graphs. The selection of the significant factors was then done by fitting a logistic regression model. This analysis was done using the R-studio statistical software.

III. RESULTS AND DISCUSSION

Preliminary Analysis

The results of this study revealed that 60.44% of the flights in 2017/2018 experienced delay (Figure 1), indicating that delays maybe a major problem at JKIA. This is in contrast to frequencies of flight delays in the developed countries, e.g., the percentages of the delayed flights in the USA in 2016 and 2017 were 15.95% and 18.57%, respectively (Bureau of Transportation Statistics, 2018). This probably may be attributed to airport infrastructural systems to handle delays in USA as compared to Kenya (Steven& Clifford, 2007).

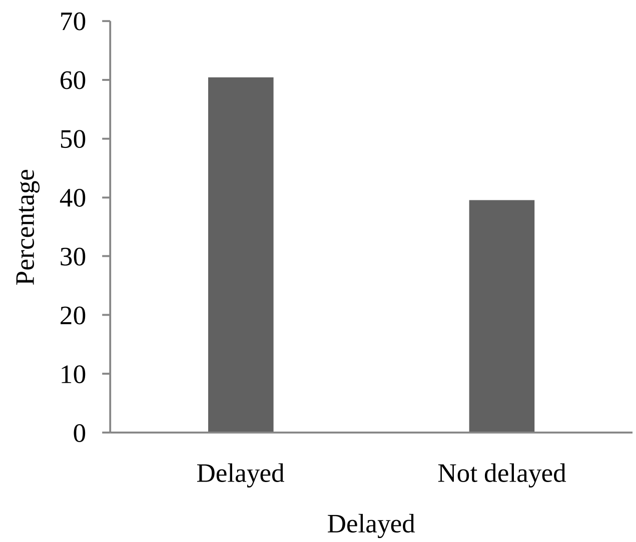

Figure 1: Summary of Flight Delays at JKIA for 2017/2018

The difference between the actual movement time and the scheduled time of aircrafts at JKIA were presented graphically (Figure 2). There were different categories of flights, i.e., (took off before the scheduled time), on schedule and delayed flights (Figure 2).Early flights occur when an airline bumps up departure times, in case everyone is on board earlier (Steve & Clifford, 2007; ICAO, 2016).

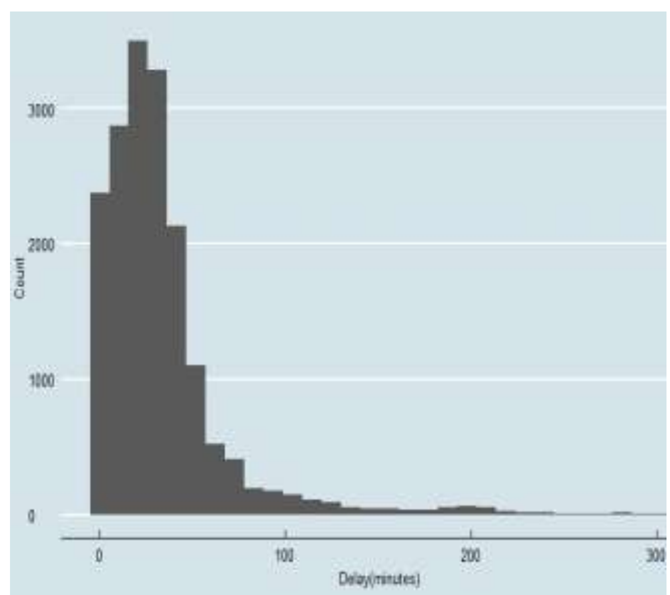

Figure 2: Distribution of Delays Times for JKIA in 2017/2018

$\boldsymbol{X}$-axis represents time difference (actual time – scheduled time); $\boldsymbol{Y}$- axis represents the number of flights; Values less than $\mathbf{0}$ on the $\boldsymbol{X}$ – axis indicate the number of minutes that a flight took off before the scheduled time.

Factors that Influence Flight Delays at Jomo Kenyatta International Airport

Flight Delays Across Airlines

The percentages of flight delays across the airlines differed, with different airlines showing different patterns of delays (Table 1). Some airlines had higher percentages of delayed flights than non-delayed flights while others had higher percentages of non-delayed flights than the delayed flights. The difference in delays may be associated with variation in management among different airlines (Ball *et al.*, 2010). In terms of individual airplanes, some showed higher frequency of delays than others. The study has attributed this to probable mechanical problems especially with age of aircraft (Rebollo&Balakrishnan, 2014). The results demonstrate that airlines as well as individual airplane should be considered while modelling flight delays.

Table 1: Percentage of Delays per JKIA Airline in 2017/2018

Airline	Delayed		Not Delayed	
	Frequency	Percentage	Frequency	Percentage
03R	2	100	0	0
05U	32	12.45	225	87.55
3J	113	66.86	56	33.14
3W	0	0	4	100
5H	2073	56.55	1593	43.45
AT	1	50	1	50
AT1	40	51.28	38	48.72
B5	442	55.88	349	44.12
BA	122	51.05	117	48.95
CZ	71	88.75	9	11.25
D3	0	0	2	100
DO	38	35.19	70	64.81
EK	546	87.64	77	12.36
ET	231	52.74	207	47.26
EY	170	25.95	485	74.05
F8	71	56.35	55	43.65
G9	266	89.26	32	10.74
JM	920	68.5	423	31.5
KL	257	92.11	22	7.89
KQ	4364	68.32	2024	31.68
LH	87	43.07	115	56.93
LX	216	97.74	5	2.26
MK	71	33.97	138	66.03
MS	17	100	0	0
PW	347	49.15	359	50.85
QR	119	19.54	490	80.46
SA	343	81.47	78	18.53
SV	177	52.37	161	47.63
TK	125	44.33	157	55.67
TM	8	17.02	39	82.98
UNS	30	44.12	38	55.88
WB	555	54.79	458	45.21
WY	105	66.04	54	33.96
XU	150	77.35	44	22.68

Flight Delays per Season

The percentage of delays during winter was higher than the percentages of delays during summer, indicating weather conditions play a role in flight delays (Figure 3). Though JKIA does not experience winter, a schedule can be delayed if the route or the destination airport is experiencing winter (Liu &Cai, 2004). In some cases, more flight delays occurred during summer than during winter (FAA, 2010). This has been attributed to fog during summer, preventing parallel runways due to reduced visibility (FAA, 2010). This result indicates that season (summer or winter) is an important variable that should be considered in modelling flight delays.

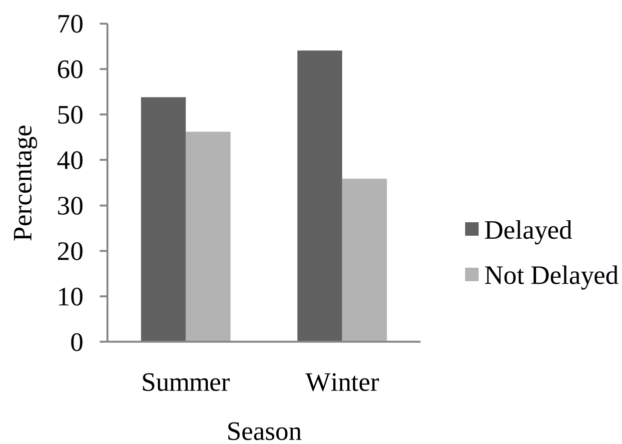

Figure 3: Flight Delays per Season for 2017/2018

Summer (March to October); Winter (October to March)

Season means summer months and winter months as provided by International Civil Aviation Organization

Distribution of Flight Delays by Month

Flight delays were variable across months (Table 2). The majority (72.7%) of the months recorded above 50% flight delays compared to about 18% non-delayed flights. The months of November to March recorded the highest number of flight delays, which coincided with the winter season. Thus, the high number of flight delays can be accounted for by poor weather conditions. Therefore, the month of flight could be an important variable to be considered in the modelling of flight delays. The findings of this study were in contrast with those of some airports in airports in USA and Europe which experience highest flight delays in June, July, August and December (Bureau of Transportation Statistics, 2015; European Union Air Transport, 2016). These flight delays patterns have been attributed to increased air traffic during those months. Therefore, traffic variability within a given airport should be considered during modelling flight delays.

Table 2: Distribution of Flight Delays by Month at JKIA for 2017/2018

Month	Delayed		Not delayed	
	Frequency	Percentage	Frequency	Percentage
January	2183	61.22	1383	38.78
February	2083	66.11	1068	33.89
March	2029	66.99	1000	33.01
April	599	54.16	507	45.84
May	598	54.46	500	45.54
June	603	54.03	513	45.97
July	787	59.22	542	40.78
August	625	50.36	616	49.64
September	520	49.48	531	50.52
October	42	40.78	61	59.22
November	2040	62.89	1204	37.199

Distribution of Delays for Domestic and International Flights at JKIA

Flight delays were also classified for both domestic and international flights (Figure 4). International flights recorded a higher percentage of delays than the domestic flights. This is contrary to findings by Arjoun*et al*. (2013) which found that in developed countries the domestic flights are delayed more than the international flights (Arjoun*et al*., 2013). This is maybe attributed to the fact that domestic aircrafts usually make more trips per day than the international aircrafts (Liu &Cai, 2004). For this reason, if an aircraft operating domestic flights suffers a significant delay, it is likely to affect its subsequent trips. The results imply that while selecting variables to be incorporated in the model for flight delays, flight types (domestic or international) need to be taken into account, with respect to region of operation.

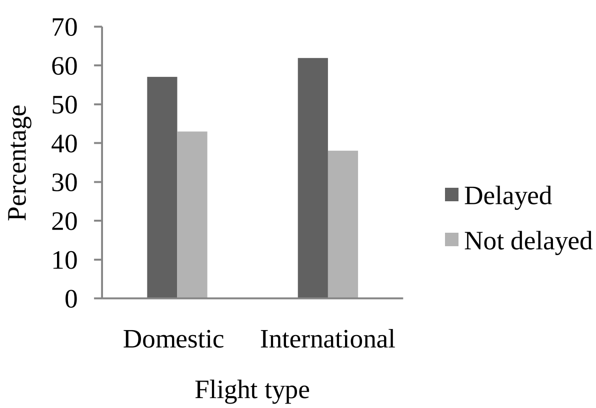

Figure 4: The Distribution of Delays for Domestic and International Flights at JKIA in 2017/2018

Flight Delays at JKIA per Day of the Week

With respect to days in a week, the highest flight delays were experienced on Fridays and Sundays, while lowest on Tuesday (Figure 5). This is in agreement with other studies that have shown that flight delays are frequent at specific day of the week (Sternberg *et al*., 2017; Bureau of Transportation Statistics, 2018). This result implies that it is important to factor in day of weeks as variable in modelling flight delays.

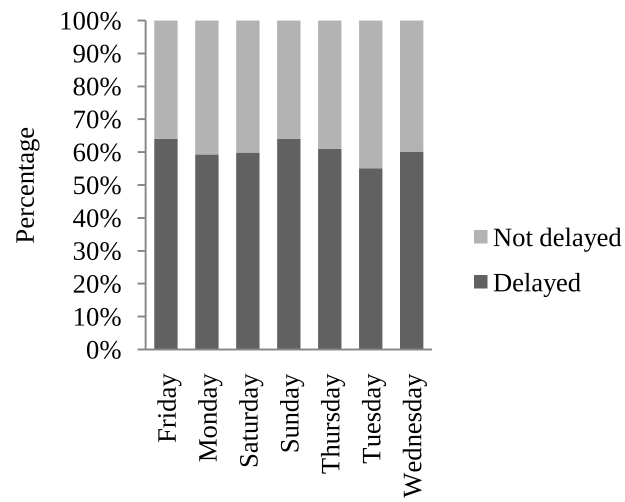

Figure 5: Flight Delays at JKIA per Day of the Week in 2017/2018

Application of Factors that Influence Flight Delays at JKIA in Developing Prediction Models

The factors that influence flight delays can be used in generating probability models that can be applied in predicting if a future flight will be on time or it will be delayed. One such model is the logistic regression model. The fitted logistic regression model was defined by its parameters, the estimates for the days of the week, months, airlines, flight class, capacity and the estimate for a night flight (Table 3). The probability values Pr (>|z|) were used to test the significance of the variables in predicting flight delays using the logistic regression model at 5% level of significance (Table 3; Table 4). The test for significance was done for each variable at a time while holding all the other variables constant. The variables that showed significance in predicting flight delays using the logistic model were flight class, the capacity of the aircraft and the months of May and July. The airlines did not show significance at 5% level of significance in predicting flight delays. In prediction of air traffic delays, significance of variables was done using probability values (p-values) (Rebollo & Balakrishnan, 2014). The significant variables in predicting the flight delays using the logistic regression model includeddays of the week, months, flight class and capacity. These are the factors that influence flight delays as selected using the the logistic regression model.

The parameters estimates were interpreted using the odds ratios. A single parameter estimate for a certain variable is interpreted when holding all the other covariates constant. For example, by holding all the other covariates constant, an international flight is $100[exp\,(0.757115) - 1] = 113.21\%$ more likely to be delayed as compared to a domestic flight. Moreover, a winter flight is $100[exp\,(0.251445) - 1] = 28.6\%$ more likely to be delayed when compared with a summer flight. A Friday flight is $100[exp(0.221661) - 1] = 24.81\%$ more likely to be delayed. An April flight is $100\,[exp(0.603842) - 1] = 82.9\%$ more likely to be delayed. A KQ flight is $100[1 - exp\,(-13.8225)] = 99.99\%$ less likely to be delayed. When logistic regression model was fitted to predict flight delays, the interpretation of the parameter estimates was also done using the signs on the values of the estimates (Cole & Donoghue, 2014). If the parameter estimate had a negative value, the interpretation was that the flight was less likely to be delayed. If the estimate was positive, then the interpretation was that the flight was more likely to be delayed (Cole & Donoghue, 2014).

Table 3: Parameter Estimates for the Fitted Logistic Regression Model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	12.69902	621.8447	0.020422	0.983707
Day Friday	0.221661	0.108412	2.044615	0.040893
Day Monday	0.014248	0.108492	0.131332	0.895513
Day Saturday	0.017456	0.108113	0.16146	0.871731
Day Sunday	0.212066	0.110886	1.912467	0.055816
Day Thursday	0.137227	0.10866	1.262907	0.206623
Day Tuesday	0.028975	0.10759	0.269308	0.787693
Month April	0.603842	0.179436	3.365218	0.000765
Month August	0.491852	0.181326	2.712522	0.006677
Month February	0.591907	0.364328	1.624656	0.104236
Month January	0.37064	0.363722	1.019018	0.308194
Month July	0.928731	0.179223	5.181986	2.20E-07
Month June	0.481102	0.180112	2.671125	0.00756
Month March	0.653066	0.349512	1.868507	0.061691
Month May	0.774802	0.182136	4.253974	2.10E-05
Month November	0.49506	0.364437	1.358422	0.17433
Month October	0.379559	0.518528	0.731993	0.464173
Airline 05U	-15.9111	621.8447	-0.02559	0.979587
Airline 3J	-14.1034	621.8447	-0.02268	0.981906
Airline 3W	-28.9023	1079.781	-0.02677	0.978646
Airline 5H	-13.3462	621.8446	-0.02146	0.982877
Airline AT1	-14.8686	621.8448	-0.02391	0.980924
Airline B5	-14.0895	621s.8446	-0.02266	0.981923
Airline BA	-15.3974	621.8447	-0.02476	0.980246
Airline CZ	-12.473	621.8451	-0.02006	0.983997
Airline D3	-29.0855	1079.781	-0.02694	0.97851
Airline EK	-13.3478	621.8447	-0.02146	0.982875
Airline ET	-14.6828	621.8447	-0.02361	0.981162
Airline EY	-15.6523	621.8447	-0.02517	0.979919
AirlineF8	-14.3135	621.8447	-0.02302	0.981636
AirlineG9	-12.2929	621.8448	-0.01977	0.984228
Airline JM	-13.1024	621.8446	-0.02107	0.98319
Airline KL	-13.3478	621.8448	-0.02146	0.982875
Airline KQ	-13.8225	621.8446	-0.02223	0.982266
Airline LH	-15.5041	621.8447	-0.02493	0.980109
Airline LX	-10.7537	621.8455	-0.01729	0.986203
Airline MK	-14.8487	621.8447	-0.02388	0.98095
Airline MS	-0.224	733.1146	-0.00031	0.999756
Airline PW	-14.5081	621.8446	-0.02333	0.981386
Airline QR	-16.1339	621.8447	-0.02595	0.979301
Airline SA	-13.2363	621.8447	-0.02129	0.983018

Table 4: Table 3 Continued

	Estimate	Std. Error	z value	Pr (>|z|)
Airline SV	-14.35	621.8447	-0.02308	0.981589
Airline TK	-14.9344	621.8447	-0.02402	0.98084
Airline TM	-28.9017	666.7596	-0.04335	0.965425
Airline UNS	-15.5863	621.8448	-0.02506	0.980003
Airline WB	-14.3916	621.8446	-0.02314	0.981536
Airline WY	-14.0838	621.8447	-0.02265	0.981931
Airline XU	-13.1359	621.8447	-0.02112	0.983147
Flight class INT	0.757115	0.098307	7.701498	1.34E-14
Capacity	0.004622	0.000809	5.711271	1.12E-08
Night	0.128341	0.199748	0.642516	0.520538
Season Winter	0.251445	0.339244	0.741194	0.458576

The fitted logistic regression model is summarised as:

ln (odds of occurrence of a delay) = 12.69902 + 0.221661 Friday + 0.014248 Monday + … + 0.028975 Tuesday + 0.603842 April + 0.491852 August + … + 0.379559 October - 15.9111 Airline 05U - 14.1034 Airline 3J + … + -13.1359 Airline XU + 0.757115 Flight class INT + 0.004622 Capacity + 0.128341 night + 0.251445 Season Winter

Perfomance of Models Developed Used Factors that Predict Flight Delays at JKIA

In a study that applied the factors that influence flight delays at JKIA, the predictive algorithms developed using machine learning approach attained an average F1 score of 76.95% (Gachoki & Muraya, 2019). This was better performance when compared to results from a study by Bandyopadhyay& Guerrero (2012) that predicted airline delays using flight departure times, and weather conditions and their prediction algorithms achieved F1 score of 56.6%. In another study predicting flight on-time performance was done using predictive features such as: the previous aircraft arriving late, weather, and departure time and achieved an average F1 score of 58.7%. (Arjoun*et al*., 2013). This shows that the factors that influence flight delays at JKIA improves the performance of prediction of flight delays.

Model Application in Prediction of Flight Delays

Once the factors that influence flight delays at JKIA have been used to develop a prediction model, it is important to show how this model can be used to predict if a future flight will be delayed or not. Prediction using logistic regression model is done using the following form of the fitted logistic regression model:

$$\hat{p} = \frac{\exp(12.69902 + 0.221661\ \text{Friday} + 0.014248\ \text{Monday} + \ldots + 0.251445\ \text{Winter})}{1+\exp(12.69902 + 0.221661\ \text{Friday} + 0.014248\ \text{Monday} + \ldots + 0.251445\ \text{Winter})}$$

Where $\hat{p}$ is the predicted probability whether a flight will be delayed or not.

For instance, prediction for an international flight on a Friday, in a month of April (summer), for a KQ airline, with a capacity of 100 passengers and travelling at night will be as:

$$\hat{p} = \frac{\exp(12.69902 + 0.221661 * 1 - 13.8225 * 1 + 0.603842 * 1 + 0.004622 * 100 + 0.128341 * 1)}{1+\exp(12.69902 + 0.221661 * 1 - 13.8225 * 1 + 0.603842 * 1 + 0.004622 * 100 + 0.128341 * 1)}$$

= 2.599708e-05

In the above equation, for the days of the week, one day was used as a reference. If a delay a flight happens on the reference day, in the equation, the day is assigned a value 0. If it happens on any other day, the day is assigned a value 1. The same case applies to the months.

Since $\hat{p}$ is less 0.5, the prediction is that the flight will be on time (not delayed).

IV. CONCLUSION

The aim of this study was to investigate the factors that influence flight delays at JKIA using predictive modelling. These are factors that can be considered when developing prediction models for flight delays. Some of the features that can be considered when developing a model to predict flight delays aredays of the week, months, flight class and capacity. Application of these factors in predictive modelling improves the performance of the prediction algorithms.

REFERENCES

[1]. Arjoun, M., Aaron, N. & Kenny, N. (2013). *Predicting Flight on-Time Performance*.Retrieved from https://pdfs.semanticscholar.org/ad5f/f1f4170218ac8b816e940e75a5e5f941fd42.pdf.

[2]. Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A. &Britto, R. (2010). *Total delay Impact Study. In Nextor Research Symposium*, Washington DC.Retrieved from http://www. nextor.org.

[3]. Bandyopadhyay, R. J. & Guerrero, R. (2012). *Predicting Airline Delays*.Retieved fromhttp://cs229.stanford.edu/proj2012/bandyopadhyayguerrero-predictingflightdelays. pdf.

[4]. Bureau of Transportation Statistics (2015). *Transportation Statistics Annual Report*. Retrieved from https://www.chegg.com/.../bureau-transportation-statistics-reports-time-performance.

[5]. Bureau of Transportation Statistics, (2018). *On-time Performance-Flight Delays at a Glance*. Retrieved from http://libguides.utoledo.edu/Statistics

[6]. Burgauer, D.& Peters, P. (2000). Airline Flight Delays and Flight Schedule Padding.*Science Investigative Report*, University of Pennsylvania, Philadephia.

[7]. Cole, S. &Donoghue, T. (2014). *Predicting Departure Delays of US Domestic Flights.*Retrieved from https://cseweb.ucsd.edu/classes/wi17/cse258-a/reports/a062.pdf

[8]. European Union Air Transport (2016). *Air Transport Industry Analysis Report.*Retrieved from http://www.academia.edu/download/3446218/Samenvatting_proefschrift_Burghouwt.doc

[9]. Federal Aviation Administration (FAA) (2010). *The Economic Impact of Civil Aviation on the U.S Economy*.Retrieved from https://www.ncbi.nlm.nih. gov/pmc/articles/PMC1117774.

[10]. Gachoki, P. K., &Muraya, M. M. (2019). Comparison of Models Used to Predict Flight Delays at Jomo Kenyatta International Airport. *Asian Journal of Probability and Statistics*, 1-8.

[11]. Hsiao, C. Y.& Hansen, M. (2006). Econometric Analysis of US Airline Flight Delays with time-of-day Effects. Transportation Research Record: *Journal of the Transportation Research Board*, 104-112.

[12]. International Civil Aviation Organization (ICAO) (2016). *Annex 14 Volume 1, Aerondromes Design and Operations.*Retrieved from www.ssd.dhmi.gov.tr/ getBinaryFile.aspx?Type=3&dosyaID=920

[13]. Ison, D. C., Weiland, L., McAndrew, I. & Moran, K. (2015). Identification of Air Traffic Management Principles Influential in the Development of an Airport Arrival Delay Prediction Model. *Journal of Aviation/Aerospace Education and Research*, 24(2), 39.

[14]. Kalliguddi, A.M.&Leboulluec AK (2017). Predictive Modeling of Aircraft Flight Delay. *UniversalJournal of Management* 5(10): 485-491.

[15]. Lawson, D., & Castillo, W. (2012). *Predicting flight delays*. Technical report, Computer Science Department, CS 229, Stanford University, Stanford, CA.

[16]. Liu, W. &Cai, X. (2004). Statistical Analysis of Airport Network of China. *Physical ReviewE*69, 046106

[17]. Mueller, E.R.&Chatterji.G.B.(2002). Analysisof Aircraft Arrival andDepartureDelayEquilibrium.*University of California, Berkeley*.

[18]. Rebollo, J. J., &Balakrishnan, H. (2014). Characterization and Prediction of Air Traffic Delays. *Transportation Research Part C: Emerging Technologies*, 44, 231-241.

[19]. Sridhar, B., Wang, Y., Klein, A., &Jehlen, R. (2009, June). Modeling flight delays and cancellations at the national, regional and airport levels in the United States. In *8th USA/Europe ATM R&D Seminar, Napa, California (USA)*.

[20]. Sternberg, A., Soares, J., Carvalho, D. & Ogasawara, E. (2017). A Review on Flight Delay Prediction.*arXiv preprint arXiv*:1703.06118.

[21]. Steven, A.& Clifford, W. (2007). The Effect of FAA Expenditures on Air Travel Delays. *Journal of Urban Economics* 63 669–678

[22]. Zou, B. & Hansen, M. (2012). Impact of Operational Performance on Air Carrier Cost Structure:*Evidence from US airlines. Transportation Research Part E: Logistics and Transportation Review*, 48(5), 1032-1048