ISSN: 2583-9292 (ONLINE)
DOI: 10.xx/ijdsml.2023.xx



# PREDICTION ANALYTICS MODEL USING CONTEMPORARY PROPOSED P-FAP AND ANALYSIS WITH ML TECHNIQUES

**R. Suguna and R. Uma Rani**
*Department of Computer Science, Sri Sarada College for Women, India*

*Abstract*

***Prediction is the essential approach for countless problems. Prediction is done for a long time by using computational techniques. Analytics is an elegant way to make predictions by using machine learning techniques. Regression is the foremost technique in a lot of problems. FAP (finite arithmetic progression) is a simple mathematical technique that is rarely used. In this research work, standard regression techniques were applied to different kinds of data, such as autism, rainfall, and child abuse data. Evaluation methods applied for the P-FAP and accuracy compared with the existing techniques Mathematical methodologies are helpful for the machine learning region. The proposed technique of P-FAP (Prediction using Finite Arithmetic Progression) is explored and satisfied results are compared with the existing methodologies.***

***Keywords:***

***Prediction Analysis, P-FAP, ML, Prediction***

## 1. INTRODUCTION

Being conscious that three different forms of data are used in this study. In addition to the category data, class labels are also contained in the autism data collection. Quantitative data comprise datasets on child abuse and precipitation. In this chapter, techniques are differentiated according to the data for each set of data. A class or category is identified from a collection of categories using classification models. That is, the algorithm can identify the group that a new variable falls into given a set of known variables. Binary predictions are frequently produced by classification algorithms used in predictive analytics. Both models of regression as well as classification are crucial components of predictive analytics. The integrity of the data utilized for the forecasts is frequently improved and cleaned using predictive modelling. To provide a more precise estimate, forecasting makes sure that the system can absorb more data, including from activities that interact with customers [1].

### 1.1 IMPORTANCE OF PREDICTIVE ANALYTICS

The analysis of historical data to help in future decision-making is known as predictive analytics. Predictive analytics can be used for a variety of data kinds, encompassing quantitative as well as qualitative information. Predictive analytics involves forecasting future behaviours, events, and results. It analyses both recent and old data using statistical methods, such as sophisticated predictive modelling and machine learning algorithms. Making accurate predictions is crucial to assisting decision-makers in navigating a world where market volatility and rapid change are constants. Predictive analyses have a lot of valuable entities when it comes to making predictions. When models are employed, cost savings are greatly impacted. Prediction analysis is used for the protection of data, to boost revenue, and to enhance productivity. Professionals can now use massive data sets to perform intricate analysis, make assumptions, retest, and then re-evaluate a model without the need for further programming thanks to developments in technologies including machine learning [1].

Effective performance in data analytics is being discovered using machine learning approaches. Data analytics is a subset of data science, which combines computer science, mathematics, and statistics. Analytical techniques come in a variety of forms, including descriptive, predictive, diagnostic, and prescriptive. Data analytics has the ability to provide a successful answer for the majority of experiments. Predictive analytics is the study of past data that aids in making effective judgments. in the future. Different types of data, including both qualitative and quantitative data, can be used with predictive analytics. The fundamental approaches advised for prediction include the ARIMA (Auto Regressive Integrated Moving Average) model, regression techniques, ANFIS (Adaptive Neuro-fuzzy Inference System), SVM (Support Vector Machine), and MLP (Multi-Layer Perceptron). Especially for categorical results, supervised machine learning techniques can be applied, and for continuous data, regression techniques can be used to handle the data [2].

### 1.2 TECHNIQUES FOR PREDICTIONS

Numerous data analysis methods, including statistics, data mining, and machine learning, are combined in predictive analytics. Since machine learning is the foundation of predictive analytics, concentrate on applying particular prediction-based techniques from the machine learning domain to enhance the understanding of upcoming events and patterns. The fundamental techniques for suggested prediction include several processes such as the ARIMA (Auto Regressive Integrated Moving Average) model, regression techniques, ANFIS (Adaptive Neuro-fuzzy inference system), SVM (Support vector machine), and MLP (Multi-Layer Perceptron [3]. There are a few popular kinds of machine learning algorithms. Neural networks, decision trees, regression analysis, and classification analysis are among them. Predictive analytics applies to models based on regression and classification in many different contexts.

### 1.3 MODELS FOR REGRESSION

Probably one of the most popular statistical techniques for research is regression analysis. In regression analysis, the action is the independent variable and the consequences are the dependent variable. The study also evaluates the degree of correlation between the two variables. An entity can use a regression method to forecast a numerical number. Stated differently, its goal is to determine the existence and strength of a relationship between variables.

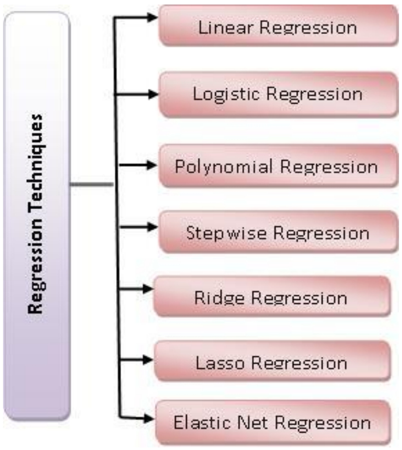


Fig.1. Types of Regressions

- Linear regression: This is the process of establishing a causal connection between the variables that are both independent and dependent. The foundation of linear regression is the conditional probability distribution.
- Polynomial Regression: When the power of an independent variable exceeds 1, curvilinear data is frequently analysed using polynomial regression. The best-fit line in the aforementioned regression technique is never a "straight line," but rather a "curve line" that fits into all of the data points.
- Stepwise regression approach: This method fits regression models with analytic representation. Several techniques are included in it, including bidirectional removal, forward decision-making and reverse elimination. Compared to other methods, stepwise regression will give access to more powerful data. When dealing with a large number of independent variables, it performs admirably.
- Ridge Regression: The analysis of multi co-linearity data that is, data in which the independent variables are highly correlated—is done using ridge regression, which is based on the ordinary least squares method. The model's standard errors are decreased by the application of ridge regression. It works with data from several regressions.
- Lasso regression: This technique aids in improving the linear model's accuracy and reducing its inconsistent behaviour. It reduces each coefficient to 0 and is mostly utilized in feature selection.
- Elastic–Net Regression: The ridge and lasso methods are combined to create Elastic-Net regression. When using connected multiple features, it is beneficial. When there are significantly more predictors than observations, applying the elastic net model of regression should be considered [4].

Autism data is suitable for classification in this study. The target parameter of the autism dataset used in this study is categorical. The majority of characteristics including gender, class variable autism, jaundice, and autistic test scores range from A1 to A10. Records including binary data, as well as age and autism outcome attributes, are numerical. The continuous variables are included in the datasets on rainfall and child abuse.

### 1.4 FINITE ARITHMETIC PROGRESSION

FAP is an effective and inherently sound mathematical method. Numerous mathematicians that are interested in research typically use and update arithmetic progression. In everyday life, development in arithmetic can be seen in things like dates, ages, days, etc. FAP is seldom ever utilized in conjunction with algorithms for computational intelligence [3].

## 2. LITERATURE SURVEY

Modular arithmetic is one of the AP concepts that were designed by Gauss, and the structure developed from it is called the clock calculator. This structure was broadly used in data encryption in 1977 by Rivest. The updated AP was derived from the sum of ratio values [4]. Arithmetic progression has been extended and applied to real-time problems. The extended part of AP was defined as a generalized arithmetic progression. Dimensional arithmetic progression was defined for two common difference series [5].Arithmetic progression effectively used for improvement of clustering centroids in the differential privacy clustering algorithm. It shows virtuous results comparing with the other techniques [6].Instead of traditional arithmetic progression, fuzzy arithmetic –geometric progressions were defined by using triangular membership value. Fuzzy values are converts into crisp value [7]. AP performed an efficient play on outlier detection along with linear regression. In this work, AP worked as a non-parametric technique and provided robust results for the data [8]. Arithmetic progression is used as a component for calculating students' creativity.

Predictive analytics has the power to make decisions and develop assumptions. Prediction techniques are applied in plenty of applications. Changes in shorelines are predicted for the Indian southeast coastal area, which is Cuddalore. Linear regression was used for the prediction of this problem [9]. Regression techniques are becoming a big part of prediction. ARIMA is another important methodology for historical data prediction. Prediction and opinion mining were done using machine learning methods. Customer health, social performance, review, and market were the components of the sentiment analysis [9].

A review was written about predictive analytics, which is the root of classical statistics. Detection of fraud, reduction of risks, and clinical choice provision systems are the main applications of prediction analytics. Prediction gives future perceptions based on past data, and there are enormous methods that can be applied to various problems of data [10].

## 3. METHODOLOGY

Workflow of the study is,

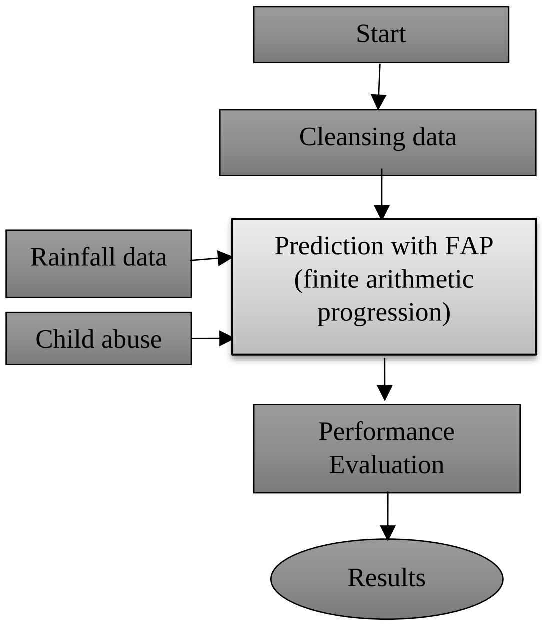


Fig.2. Flow of the work

### 3.1 DESCRIPTION OF THREE KINDS OF DATA

There are 292 cases and 21 attributes in the autism data. Information on the children includes age, gender, ethnicity, jaundice at birth, family members who have PDD, country of residence, if they have previously used the screening app, the type of screening method, and 10 screening method-related questions with yes/no responses. The UCI repository is where these data are downloaded from. There are 115 instances and 16 properties in rainfall data. This information was taken from the Indian government's agriculture website. The data used from NCRB on child abuse consists of 14 attributes and 13 incidents.

### 3.2 PRE-PROCESSING

A type of data understanding is data pre-processing. It comprises data cleaning, which entails the management of missing numbers, the removal of outliers, and the improvement of data quality. In this paper, missing values were handled with mean for the autism, rainfall data and child abuse data.

### 3.3 PROPOSED TECHNIQUE FOR PREDICTION (P-FAP)

#### *3.3.1 Finite Arithmetic Progression:*

FAP is a mathematical technique that is both successful and fundamentally viable. Arithmetic progression is essentially found in everyday situations, such as dates, ages, and days. Seldom is FAP utilized in conjunction with algorithms for computational intelligence. Arithmetic progression was often used and updated for research by many ambitious mathematicians. For many issues, the majority of basic mathematical notions provide a great deal of satisfaction. The study of mathematical series advances our understanding of computation and probability by facilitating models and predictions for a range of performances. Because rigorous mathematics reduces operating time and energy costs, their research enables the creation of a reliable statistical forecast. A genuine series could have additional influencing variables that, when the simulation is built, are crucial to the information generated, necessitating the creation of a historical data structure that can incorporate these particularities (Gleison et al, 2020). In this study, data on child abuse and rainfall forecasts are made using finite arithmetic progression. Arithmetic FAP in its general form is,

$$a_n=a+(n-1)d \tag{1}$$

Difference d is calculated as,

$$d=a_i-a_j=d_1-a_k=d_2-a_l\ldots.a_{n-1} \tag{2}$$

Data on child abuse and rainfall are not always the same. Due to the possibility that some years would have more or less rainfall and child abuse may also fluctuate in the same way. Thus, the d value is determined by taking the differences of the subsequent values. The following Fig.3 shows that the workflow of the proposed technique P-FAP is,

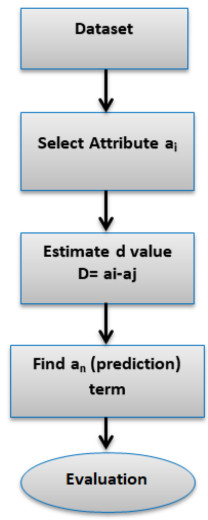


Fig.2. Flow of proposed model

The proposed model FAP (finite arithmetic progression is applied for rainfall and child abuse data. The following is the pseudo-code for making predictions using the finite arithmetic progression (FAP):

*P-FAP*
*Step1: Select Attribute* $a_i$
*Step2: Identify n value*
*Step3: Calculate d value*
$d \rightarrow a_i-a_j-a_k\ldots\ldots a_n$
$d1 \rightarrow a_i-a_j$
$d2 \rightarrow a_j-a_k..$
$d_n \rightarrow a_{n-1}-a_n$

*d->d1-d2…dn*
*Step 4: Find the n+1$^{th}$, n+2$^{th}$ term.*
*Step 5: end*

# 4. RESULTS AND DISCUSSION

## 4.1 CLASSIFICATION FOR AUTISM DATA

For binary classification, the classification and regression techniques that are maintained are J48, KNN, SVM, naïve Bayes, decision trees, neural networks, and logistic regression. Among the other algorithms, KNN has 99.8% accuracy for autism data. The Fig.4 shows the accuracy of classification algorithms.

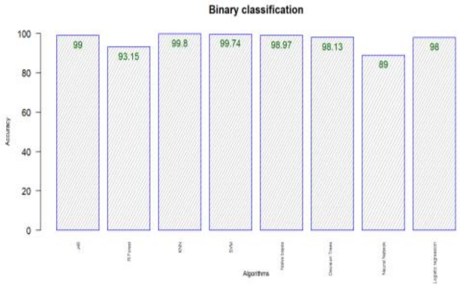


Fig.4. Accuracy comparison for Autism data

## 4.2 RAINFALL DATA

### *4.2.1 Multiple Linear Regression on Rainfall Data:*

```
Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept) -0.0548132  0.1149990   -0.477    0.635
JAN          0.9991073  0.0010190  980.506   <2e-16 ***
FEB          0.9991680  0.0008826 1132.029   <2e-16 ***
MAR          1.0009441  0.0008253 1212.884   <2e-16 ***
APR          0.9999059  0.0009800 1020.322   <2e-16 ***
MAY          1.0001094  0.0006625 1509.546   <2e-16 ***
JUN          1.0001078  0.0002923 3421.491   <2e-16 ***
JUL          1.0001807  0.0002610 3831.835   <2e-16 ***
AUG          1.0001001  0.0002990 3345.143   <2e-16 ***
SEP          0.9998956  0.0003213 3111.854   <2e-16 ***
OCT          0.9996650  0.0003969 2518.792   <2e-16 ***
NOV          1.0001170  0.0006158 1623.986   <2e-16 ***
DEC          1.0008260  0.0011643  859.563   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1035 on 102 degrees of freedom
Multiple R-squared:      1,    Adjusted R-squared:      1
F-statistic: 1.086e+07 on 12 and 102 DF,  p-value: < 2.2e-16
```

Fig.5. Residual summary of rainfall data

The rainfall data's multiple linear regression result is examined in Fig.5. R-squared is given as 1. As a result, the model fits rainfall data well. The majority of the variables are significant for the model, with a p-value of less than 0.05.

### *4.2.2 Prediction using Proposed Technique P-FAP on Rainfall Data:*

The Fig.6 examines quality measurements like recall, precision, and f1 scores that are close to 1, as well as error metrics like MAE, MSE, and RMSE that are close to 0. Thus, using autism data, the algorithms worked well. Table 1 illustrates that rainfall forecasting with both moderate and high-level estimates is from 2016 to 2025. The confidence intervals are 95% and 80%. Fig.6 shows the forecasting process in graphical form.

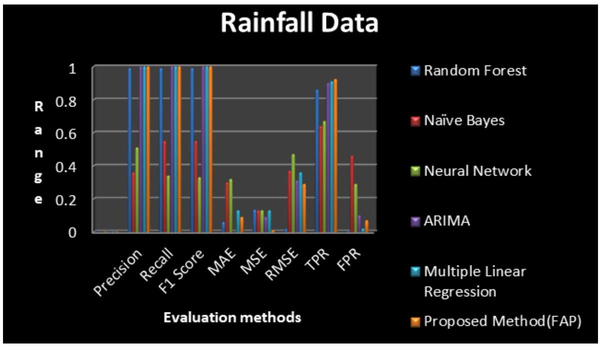


Fig.6. Comparison of P-FAP with existing techniques

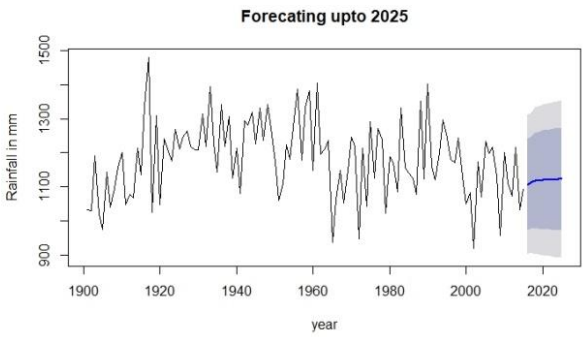


Fig.7. Prediction up to 2025 year

Table.1. Prediction values for rainfall using P-FAP

| Year | Low 80 | High 80 | Low 95 | High 95 |
|---|---|---|---|---|
| 2016 | 969.6137 | 1237.231 | 898.7797 | 1308.065 |
| 2017 | 979.6177 | 1248.235 | 908.5191 | 1319.333 |
| 2018 | 974.9193 | 1256.383 | 900.4205 | 1330.881 |
| 2019 | 976.0551 | 1259.546 | 901.0196 | 1334.581 |
| 2020 | 974.9715 | 1262.418 | 898.8889 | 1338.501 |
| 2021 | 975.6334 | 1265.474 | 898.9174 | 1342.190 |
| 2022 | 972.2957 | 1264.912 | 894.8448 | 1342.363 |
| 2023 | 971.8897 | 1266.972 | 893.7861 | 1345.076 |
| 2024 | 970.9258 | 1268.541 | 892.1519 | 1347.315 |
| 2025 | 968.5466 | 1268.615 | 889.1232 | 1348.039 |

### *4.2.3 Linear Regression on Child Abuse Data:*

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  105.18867  132.06988   0.796    0.426
crime$`2012`   6.28902    0.04243 148.220   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2895 on 492 degrees of freedom
Multiple R-squared:  0.9781,  Adjusted R-squared:  0.9781
F-statistic: 2.197e+04 on 1 and 492 DF,  p-value: < 2.2e-16
```

Fig.7. Linear regression of child abuse data

The Fig.7 displays the linear model's residual summary for data on child abuse. Both the predictor's and the model's p values are below the significance threshold. Higher values are expected for R-squared and Adjusted R-squared. The model's performance on the data was good, yielding 0.9781. The data correlate almost 1. The data thus shows a positive correlation.

#### *4.2.4 Prediction using P-FAP on Child Abuse Data:*

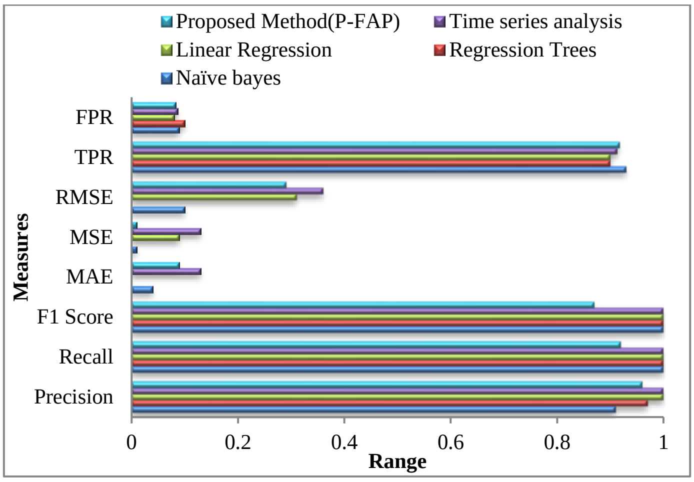


Fig.8. Evaluation metrics for child abuse data

Measures of evaluation for child abuse data are explored in Fig.8. More accurate predictions are produced using efficient current prediction algorithms in the data on child abuse. In comparison to the current approaches, the P-FAP technique that was proposed yielded satisfactory outcomes. The estimated values for the suggested approach P-FAP are displayed in Table 2 below.

Table.2. Prediction values for child abusing using P-FAP

| Year | CI:80% | CI:95% |
|---|---|---|
| **2021** | 42922.8 | 40233.1 |
| **2022** | 31051.8 | 36467.9 |
| **2023** | 40231.6 | 39605.6 |
| **2024** | 696495.9 | 788619.4 |
| **2025** | 193960.5 | 196056.1 |

## 5. CONCLUSION

In computational intelligence approaches, arithmetic progression is an antiquated method that is sporadically employed. This chapter applies FAP to the predictive analytics methods processed on child abuse and rainfall data. When compared to the current procedures, it yields better outcomes. Future improvements to FAP will raise its accuracy level and produce more effective predictions. P-FAP provided satisfactory results for rainfall data compared with the child abuse data. And prescriptive analytics takes over to optimize the methods applied to the aforementioned data that are analysed.

## REFERENCES


[1] Wael Said, "Data Mining Techniques for Database Prediction: Starting Point", *Journal of Theoretical and Applied Information Technology*, Vol. 65, No. 2, pp. 1-13, 2018.

[2] Dharmendra Kumar Yadav, "General Study on Two-Dimensional Generalized Arithmetic Progression", *Acta Ciencia Indica*, Vol. 35, No. 2, pp. 557-566, 2008.

[3] Gleison Guardia and Nayara Bonim Do Nascimento, "Arithmetic Sequence with Multiple Reasons", *American Journal of Engineering Research*, Vol. 7, No. 5, pp. 370-376, 2017.

[4] Zexuan Fan and Xiaolong Xu, "APDPk-Means: A New Differential Privacy Clustering Algorithm based on Arithmetic Progression Privacy Budget Allocation", *Proceedings of IEEE International Conference on Data Science and Systems*, pp. 1-6, 2019.

[5] Kirsten Martin, "Predatory Predictions and the Ethics of Predictive Analytics", *Journal of the Association for Information Science and Technology*, Vol. 74, No. 5, pp. 531-545, 2023

[6] Ezzatallah Baloui Jamkhaneh, "The Progressions of Fuzzy Numbers and Their Features", *Journal of Mathematics and Computer Science*, Vol. 2, No. 1, pp. 20-26, 2011.

[7] Logesh Natarajan, "Shoreline Changes Over Last Five Decades and Predictions for 2030 and 2040: A Case Study from Cuddalore, Southeast Coast of India", Springer, 2021.

[8] M.A. Hussein, "Outlier Detection Method in Linear Regression Based on Sum of Arithmetic Progression", *The Scientific World*, Vol. 2014, pp. 1-13, 2014.

[9] Gaurav Gosavi, "Prediction Techniques for Data Mining", Proceedings of National Conference on Emerging Trends in Computer Engineering and Technology, pp. 1-7, 2022.

[10] Sri Redjeki, "Big Data Analytics for Prediction using Sentiment Analysis Approach", *Journal of Theoretical and Applied Information Technology*, Vol. 100, No. 13, pp.1-12, 2021.

[11] Vaibhav Kumar and M.L. Garg, "Predictive Analytics: A Review of Trends and Techniques", *International Journal of Computer Applications*, Vol. 182, No. 1, pp. 1-12, 2018.

[12] M. Rustagi and N. Goel, "Predictive Analytics: A Study of its Advantages and Applications", *International Research Journal*, Vol. 12, No. 1, pp. 1-14, 2022.

[13] M Tohir, "Students Creative Thinking Skills in Solving Two Dimensional Arithmetic Series through Research-Based Learning", *Journal of Physics*, Vol. 108, pp. 1-10, 2018.

[14] S. Feng, C. Hategeka and K.A. Grepin, "Addressing Missing Values in Routine Health Information System Data: An Evaluation of Imputation Methods using Data from the Democratic Republic of the Congo during the COVID-19 Pandemic", *Population Health Metrics*, Vol. 19, pp. 44-54, 2021.

[15] Fadi Thabtah, "An Accessible and Efficient Autism Screening Method for Behavioural Data and Predictive Analyses", *Health Informatics Journal*, Vol. 44, pp. 1-17, 2018.

[16] I.D. Dinov and M. Darcy, "Predictive Big Data Analytics: A Study of Parkinson's Disease using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations", *PLoS ONE*, Vol. 11, No. 8, pp. 1-11, 2016.