Sabyasachi Mohanty *et al*, International Journal of Computer Science and Mobile Applications,
Vol.5 Issue. 10, October- 2017, pg. 48-57 **ISSN: 2321-8363**
Impact Factor: 5.515

GovSchemAna - A Machine Learning Enabled Android App for Analysis of Government Schemes

Sabyasachi Mohanty[1], Sudarsan Padhy[2], M Vamsi Krishna[3]

[1]*Department of Computer Science & Engineering, Centurion University of Technology & Management, Bhubaneswar, Odisha, India, Email: pikunimohanty@gmail.com*
[2]*Department of Computer Science & Engineering, Silicon Institute of Technology, Bhubaneswar, Odisha, India, Email: spadhy07@gmail.com*
[3]*Department of Computer Science & Engineering, Centurion University of Technology & Management, Paralakhemundi, Odisha, India, Email: vamsimangalam@gmail.com*

Abstract

Smartphones and machine learning as part of today's technical world have demonstrated their benefits in their respective fields of operation. Smartphones have become the necessity of every human being that functions more than just as a communicating device. Though these mobile devices are designed to operate in less processing power, storage, and energy, these are sufficient to run the variety of apps such as calculator, watch, alarm clocks, and camera. Similarly, machine learning facilitates systems that can learn from experience and perform analytics. Advances in machine learning have greatly supported the decision making ability, though a sophisticated platform is required to execute the algorithms. Considering the limitations of both the technologies i.e., machine learning and smartphone, in terms of computing power, memory, and energy, we propose a novel approach for performing analysis of government schemes using a machine learning enabled Android app. This app, named as GovSchemAna (Government Scheme Analyzer), offers an on-device platform to perform predictions in the field of government schemes for their stakeholders.

***Keywords**:* Machine Learning, Android, Government Scheme.

1. Introduction

Welfare and developmental schemes are offered by government for the social and economic upliftment of the society (Saxena, 2007; Ministry of Electronics & Information Technology, 2017). In order to achieve maximum benefit through the collective performance of the schemes, every aspect starting from understanding the problem area to planning and implementation to monitoring and assessment has to be flawlessly executed. But in a democratic country like India, which has different forms of government at different levels, addressing the integral undesired phenomena of our contemporary society such as poverty, illiteracy, crime, racism, unemployment, and environmental pollution through schemes is no easy task. So, planning and analysis of the schemes that go hand in hand and also lead to strategies to receive desired outcomes with reduced risk, is given utmost importance. To address this complex and challenging task all the stakeholders should have insight and futuristic views to attain maximum benefit. This aspect can be greatly supported by machine learning (Alpaydin, 2014; Haykin, 2009). It is basically, programming machines to optimize a performance criterion using past experience or data samples (Alpaydin, 2014). Today scheme based data are available to the public (Ministry of Electronics & Information Technology, 2017), but literally speaking no avenue is available for analyzing the data with ease for them.

The central problem of poverty is a challenging issue of rural development that has been greatly benefited from the implementation of decentralized planning which involves people at the grassroots level (Sahota, 2008; Ministry of Rural Development, 2017). In India, there are two and a half lakh Gram Panchayats and their respective Gram Sabhas or village meetings plan for the developmental activities of their respective localities. This structure has benefited with better results as members of a Gram Sabha better understand the ground reality of its place, but sometimes fail to accurately forecast the influential factors. This issue of poverty has been greatly supported by MGNREGA (Mahatma Gandhi National Rural Employment Guarantee Act) scheme (Satish *et al.*, 2013; Das, 2013; Ministry of Rural Development, 2017) which aims to enhance the livelihood security of people in rural areas and to boost the rural economy. It guarantees the 'right to work' by providing a minimum of 100 days of wage employment to members of rural households who have volunteered to do unskilled manual work (Das, 2013). So it has to be planned in such a way that, there is sufficient unskilled manual work to fulfill its demand. As the strategy behind follows a bottom-up approach, an analysis is to be done at the ground level in advance to get the projects sanctioned and support the work demand. Therefore forecasting the upcoming work demand of a locality is a critical problem before the stakeholders, which includes the members of the Gram Sabhas as well as the government officials. The Government of India (GoI) has made MGNREGA scheme related information available to the public on the web portal http://www.nrega.nic.in/ so that the stakeholders are aware of the scenarios (Ministry of Rural Development, 2017). Unfortunately, there is no avenue for the stakeholders to foresee or analyze the work demand at the Gram Panchayat or Block level. Therefore it is highly essential to provide a platform to compare and analyze the scenarios, which satisfies the needs of the stakeholders.

Gross Value Added (GVA) at factor cost as a reflection of country's productivity is highly influenced by the government schemes. In India, schemes are offered by the central government as well as governments at states or union territories. With the responsibility to support all the varieties of schemes, Government of India (GoI) allocates budget for the central plans, and states and union territories plans. Mohanty and Padhy (2017) performed a similar kind of study using OFS-TLBO-SVR hybrid machine learning model for optimal budget allocation of government schemes to maximize GVA at factor cost (Mohanty and Padhy, 2017). For the government administrators and financial advisors, GVA prediction based on budget allocation to the schemes at center and states is a critical and challenging task. There are many such scheme related scenarios that require a predictive system in hand.

Being a phenomenal invention of this age, mobile phone has become part and parcel in the life of all individuals. Advances in ICT (Information and Communication Technology) and mobile phones have empowered everyone with its computing capability, storage, sensors, and network connectivity. Nowadays, availability of low-cost smartphones have made them affordable to everyone and due to easy handling, operating smartphones have become child's play. Their increasing usability in a variety of activities involving emergency to entertainment has made them very popular. These hand held smart devices are also considered as intelligent devices because of their capability to execute specialized machine learning algorithms to produce predictive results based on the accessible information. Availability of information play a vital role in the field of analytics and internet has become an easy, quick, and low-cost mode of transferring information from one to another. Policy-wise, the government is also very promising in this aspect of sharing important information to the stakeholders through different web portals and mobile apps.

Android (Google, 2017) is the world's most popular mobile operating system and is heavily exploited by the smart devices. Today there are more than two million apps available for the Android devices (Google, 2017). Among them, there are many specialized Android applications that have incorporated machine learning algorithms. One such application is iWander (Sposaro *et al.*, 2010), which has been designed to assist dementia patients and their caregivers. It uses a Bayesian network to perform classification between normal and abnormal behaviors of patients.

In this study, we present an Android application, named as GovSchemAna (Government Scheme Analyzer), specially designed for the stakeholders of the government schemes to perform regression, classification, or time series forecasting based on the existing machine learning techniques such as Linear Regression, Multilayer Perceptron, and Support Vector Machine (SVM).

The rest of this paper is organized as follows. In Section 2, the software components used as the building blocks along with the processes involved in the Android application under study, i.e., GovSchemAna, are explained. Then in Section 3, results obtained from the experiments performed using the real scheme based data available with government are presented. The difficulties and challenges are highlighted in Section 4, and finally, the paper is concluded in Section 5.

2. Methods and Design

2.1 Software

Our application is designed on the Android platform along with a set of machine learning libraries such as Weka (Witten *et al.*, 2016; Hall *et al.*, 2009), LibSVM (Chang and Lin, 2011), and TimeSeriesForecasting (Hall, 2017).

Android is an open source mobile native software stack produced by Google (Google, 2017; Collins *et al.*, 2011; Meier, 2012; Pradhan and Deshpande, 2014). The layered architecture of the Android platform packages a Linux kernel (an operating system) with native libraries, Android runtime, application framework, and applications (Pradhan and Deshpande, 2014). The Android SDK (software development kit) provides the necessary libraries that interface with the underlying hardware to develop or deploy Android applications (Google, 2017; Collins *et al.*, 2011; Meier, 2012). Applications written in Java are designed to run on Android runtime which has Dalvik Virtual Machine (DVM) and core libraries as its key components. Android uses both flat files and SQLite database to store persistent data. It also has the ability to easily thread background running processes, and are compatible with other Android devices like Tablets, and Android TV.

Unlike dedicated systems, our application can co-exist with other installed applications on the device. GovSchemAna shares resources with other applications in such a way that it runs inconspicuously while using limited resources. From the operational point of view, firstly a scheme related dataset is created and stored on the device, and then a machine learning task is performed based on the technique chosen by the user. To keep the device responsive, an asynchronous task is created which takes all the responsibility of performing the machine learning task. Once the processing is over, the final results are displayed on the screen.

Machine learning systems fuel many aspects of modern world. Behind the scene, they are powered by machine learning libraries, which are the program implementations of their corresponding algorithms. Weka (Witten *et al.*, 2016; Hall *et al.*, 2009) is an open source data mining software that provides a comprehensive collection of machine learning algorithms and data pre-processing tools to researchers and practitioners. It contains tools for classification, regression, clustering, and association rules. This platform allows users to quickly implement and compare different state-of-the-art techniques in machine learning such as Multilayer Perceptron, Support Vector Machine, and Random Forest (Alpaydin, 2014; Haykin 2009). The algorithms can either be directly applied to a dataset or called from a Java program. From the implementation point of view, Classifier class is basically at the core of it and there are a lot of common options or parameters. All the machine learning techniques are the concrete extensions of this class.

For our application, we have used Weka-for-Android (University of Waikato, 2011) which is a stripped version of Weka. For the implementation of Support Vector Machine, LibSVM (Version 1.0.8) (Chang and Lin, 2011) has been included. TimeSeriesForecasting (Hall, 2017) library is also included to provide a time series forecasting environment for our application. Technically, it is a wrapper for a regression method that automates the process of creating lagged variables and date-derived periodic variables and provides the ability to do closed-loop forecasting (Hall, 2017).

2.2 Application Functionalities

The motive behind the development of GovSchemAna Android application is to facilitate with an analytical platform that will suffice the needs of the stakeholders of the government schemes. For smooth and easy operation the GovSchemAna app is fundamentally composed of 3 modules to handle the following important functionalities.

- Building dataset
- Initializing and processing of machine learning task
- Managing results

Figure 1 shows the screenshot of the fundamental sections represented in the menu of GovSchemAna application.

Dataset creation is an integral and critical part for the field of machine learning. High quality labeled dataset is really difficult to produce. The experience of the designer helps to identify the desired features of the samples in the dataset. So taking this aspect into consideration the user of the application is facilitated with a special functionality of developing datasets that can be used for the machine learning tasks. Though few data sources (like www.data.gov.in) related to government schemes are already available, the user is free to collect more data from any kind of sources such as data servers, files, and internet or cloud. The data collected may require to undergo pre-processing to address the issues related to missing data, normalization, scaling, transformation, etc. As the experiment may demand a single full dataset or split dataset having training and testing parts, there is an option available to the user to do the needful. Finally, the dataset is written onto the storage of the device. The data can be saved in the flat files or SQLite database. The detailed flowchart of building the dataset is shown in Figure 2.

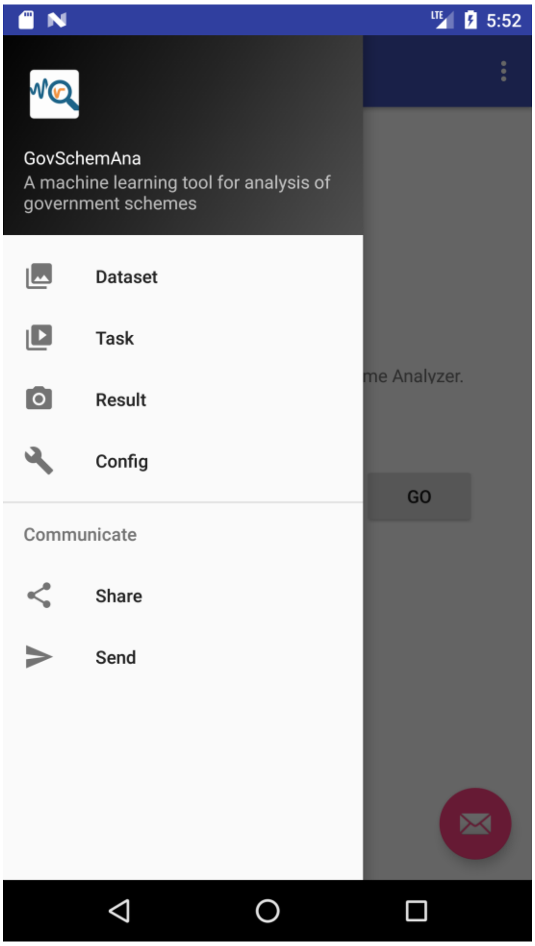

Figure 1: Screenshot of Menu

Sabyasachi Mohanty *et al*, International Journal of Computer Science and Mobile Applications,
Vol.5 Issue. 10, October- 2017, pg. 48-57 **ISSN: 2321-8363**
Impact Factor: 5.515

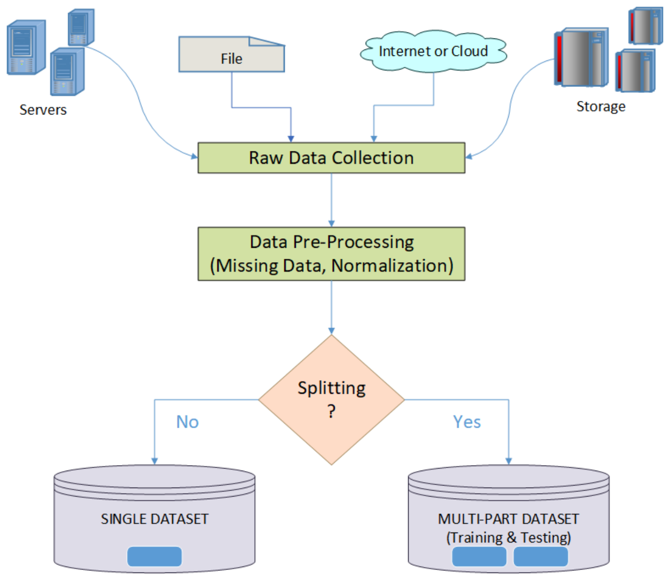

Figure 2: Flowchart for Building Dataset

Figure 3 shows the screenshot of the task entry user interface of GovSchemAna app. Here the user is allowed to select the type of the task (i.e., classification, regression, or time series forecasting), the desired dataset and the technique or machine learning algorithm to be used. The user is also facilitated to reinitialize or change the parameters or options of the selected technique. The application starts its learning process on receiving the click event from the Execute button. Technically, in the background, a new Thread of execution starts using AsyncTask to run the machine learning task asynchronously (Meier, 2012). The AsyncTask class of Android (Meier, 2012; Pradhan and Deshpande, 2014), implements a best practice pattern to ensure that the application responds quickly to the user interactions or system events by moving all processing and I/O (input/output) operations off the main application Thread and into a child Thread. This critical attribute of responsiveness for our app is attained by moving the learning task, which is a time-consuming operation, onto the background Thread and synchronizing with UI (user interface) Thread for updates. The functionality of the Abort button is to cancel the ongoing learning process by the AsyncTask. Once the model is trained, it attains the power to predict or forecast. Therefore to reuse the model at a later stage of time, Save button has been provided to save the model. The flowchart, shown in Figure 4, represents the process followed to build a machine learning model.

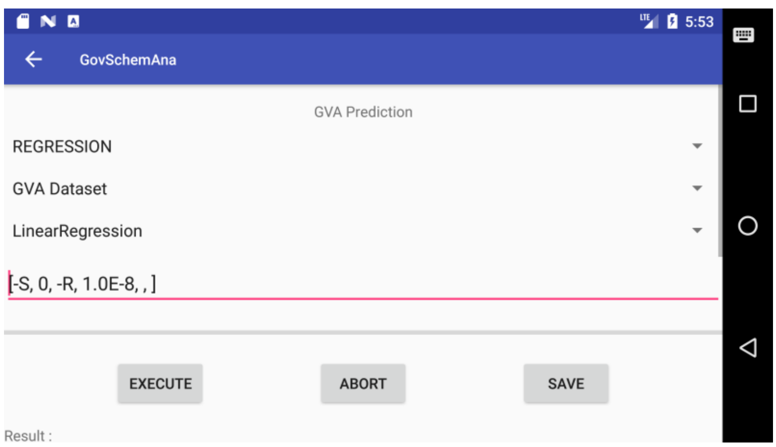

Figure 3: Screenshot of Task Entry User Interface

Sabyasachi Mohanty *et al*, International Journal of Computer Science and Mobile Applications,
Vol.5 Issue. 10, October- 2017, pg. 48-57 **ISSN: 2321-8363**
Impact Factor: 5.515

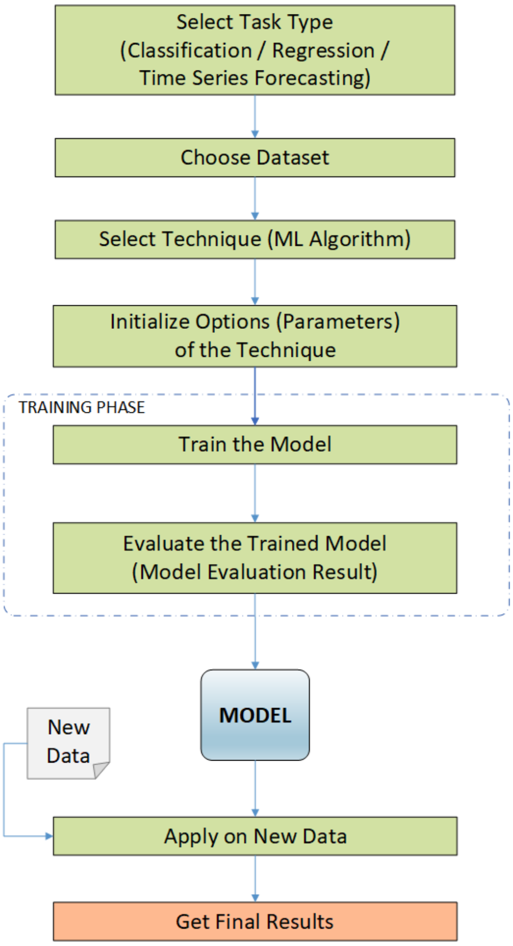

Figure 4: Flowchart for Building a Machine Learning Model

Each experiment defined as a task produces a set of results. Figure 5 shows the screenshot of the results obtained using the GovSchemAna application. Basically, it consists of the evaluation measures of the model for the configured task and a graphical representation of the prediction outcome.

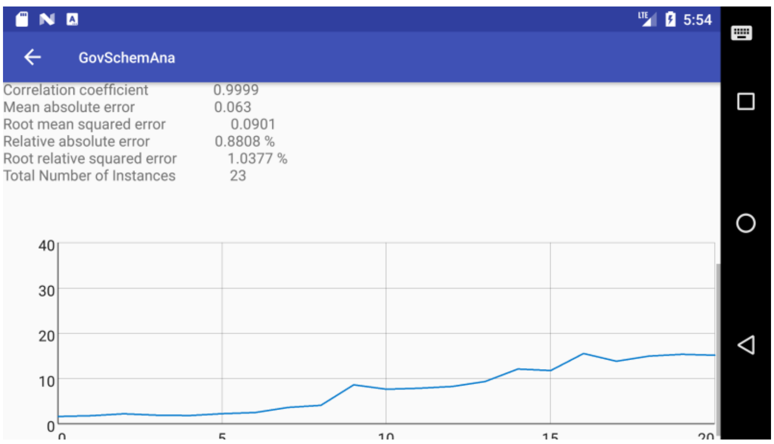

Figure 5: Screenshot of Result Screen

This application needs to be configured as per the intent of the user. So, in the configuration dashboard, the user virtually manages all aspects of the system. Necessary user information and its purpose are also required to be recorded for future references. This part of the application also facilitates customizations required to enhance the user experience and efficient use of the constrained environment that has limited muscle power in terms of processing, memory, and battery present on the device.

3. Results and Discussions

As examples of usage of our Android application, we test the models on scheme based datasets and illustrate the feasibility of the proposed approach. To evaluate the accuracy of the machine learning models available on the app, we conducted a comparative study with the desktop version of Weka (Version 3.9.0). Here we have used four standard statistical metrics. They are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE).

Forecasting of work demand under MGNREGA scheme is a critical task before the stakeholders to properly plan for the developmental projects and the course of action required to fulfill the demand of work in the desired time frame. Data related to monthly work demand depicting the number of persons requested for unskilled work is available on the web portal of MGNERGA. So for this forecasting (time series forecasting) task, a dataset is created based on the monthly work demand pattern of Basudevpur Block, under Bhadrakh District of Odisha (India), from April 2012 to March 2017. A lag of 12 months is also applied to it. For an in-depth study, we performed a comparative analysis using different machine learning techniques. Table 1 shows the comparison of experimental results based on Linear Regression, Multilayer Perceptron, and Support Vector Machine using work demand dataset for Basudevpur Block. It is observed that both the desktop Weka (Version 3.9.0) and our Android app produces equal evaluation results. In addition to this, it is also found that the Support Vector Machine method with Radial Basis Function (RBF) kernel yields higher accuracy (RMSE = 0.1) than the other two.

Table 1: Comparison of Performance of Different Techniques on Work Demand Forecasting Task

	Desktop Weka (3.9.0)				GovSchemAna			
Technique	**MAE**	**RMSE**	**RAE (%)**	**RRSE (%)**	**MAE**	**RMSE**	**RAE (%)**	**RRSE (%)**
Linear Regression	0.4746	0.6366	34.9157	29.2246	0.4746	0.6366	34.9157	29.2246
Multilayer Perceptron	0.161	0.2042	12.0893	9.4246	0.161	0.2042	12.0893	9.4246
Support Vector Machine (RBF Kernel)	0.1	0.1	7.3504	4.5667	0.1	0.1	7.3504	4.5667

To address the issue of predicting GVA at factor cost, a regression task is framed with a dataset containing 24 features that include budget support for central plan and central assistance for states and union territories plan. The dataset is built with 37 data samples representing the annual figures of immediate past 37 years (i.e., from the financial year 1980-81 to 2016-17). The comparative study using Linear Regression, Multilayer Perceptron, and Support Vector Machine shown in Table 2 confirms that both GovSchemAna and desktop Weka produces same evaluation results and Support Vector Machine with Radial Basis Function (RBF) kernel outperforms the other two methods with MAE of 0.5618 and RMSE of 0.8782. Hence the user can now derive his conclusions based on the predictive results.

Table 2: Comparison of Performance of Different Techniques for Prediction of GVA at Factor Cost

	Desktop Weka (3.9.0)				GovSchemAna			
Technique	**MAE**	**RMSE**	**RAE (%)**	**RRSE (%)**	**MAE**	**RMSE**	**RAE (%)**	**RRSE (%)**
Linear Regression	0.7417	1.0839	2.3743	2.7718	0.7417	1.0839	2.3743	2.7718
Multilayer Perceptron	1.0829	1.5392	3.4664	3.9364	1.0829	1.5392	3.4664	3.9364
Support Vector Machine (RBF Kernel)	0.5618	0.8782	1.7982	2.2459	0.5618	0.8782	1.7982	2.2459

Sabyasachi Mohanty *et al*, International Journal of Computer Science and Mobile Applications,
Vol.5 Issue. 10, October- 2017, pg. 48-57 **ISSN: 2321-8363**
Impact Factor: 5.515

E-Governance solutions also facilitate analytics to evaluate and monitor performance in the domain of schemes. IGoSA (Intelligent Government Scheme Advisor) (Mohanty *et al.*, 2015) is intending to address this aspect to produce more successful schemes by taking the help of machine learning techniques and data management solutions. As the GovSchemAna app is found to produce acceptable results, it can now greatly support and take the hassle out of such systems by offering predictive models at the finger tips. It is also observed that in addition to the benefits of mobility, the followings are some of the exciting advantages associated with this Android application.

1. GovSchemAna is a low budget and cost effective solution as investing in a similar web solution would be much more costly.
2. The user can perform a comparative study of a problem based on different techniques and store their results for future references.
3. There is no need for internet or support of external agencies to process the analytical tasks.
4. Users can perform their own predictive analysis and if required, share with others with accuracy details.

4. Challenges

Using smartphone technology for data analytics has numerous advantages in cost and capability of the system. However, leveraging an existing system does pose challenges that the user can identify and avoid.

4.1 Hardware

The Android platform, though robust, is powered by the underneath hardware configuration of the smartphone. As the processing of the machine learning algorithms is a computing intensive task, one of the biggest challenge is that the devices with low computing power or memory may fail to successfully execute the algorithms and hence cannot produce the desired analytical results. In addition to that, all mobile devices do not possess equal processing power and storage, and there are a variety of manufactures, who are producing differently configured hardware components.

In addition to this, different machine learning algorithms require a different minimum amount of processing power and memory. Therefore, situations may arise where low computation intensive algorithms like Linear Regression may succeed to get processed whereas high computation intensive algorithms don't.

4.2 Data Volume

As the primary goal of analytics is to mine enough data to build an efficient model, more data is expected to produce better results. But the mobile phones have very limited storage. Therefore it is not feasible to hold a high volume of data for processing.

4.3 Implementation

Though the concepts behind machine learning are easy to understand but require an in-depth understanding to fully grasp their benefits. There are a variety of algorithms such as Linear Regression, Support Vector Machine, and Multilayer Perceptron with different inherent concepts such as filtering, lag creation, and data normalization (Witten *et al.*, 2016). As the objective is to create an efficient model, all the configurable parameters need to be properly initialized considering the behaviour of the problem in hand. Therefore the users of the GovSchemAna app are required to fully understand the use of the algorithms before getting into their implementations.

5. Conclusions

The Android application GovSchemAna is built to provide a platform featured with machine learning for analyzing different aspects of government schemes. We tested the app with two different problems, i.e., work demand forecasting, and prediction of Gross Value Added (GVA) at factor cost using techniques such as Linear Regression, Multilayer Perceptron, and Support Vector Machine. On comparison, it is found that for both the tasks, our GovSchemAna app and the desktop version of Weka application generated same predictive results. Hence this app can be used as an alternative to machine learning desktop applications and can also support E-Governance applications having similar objectives. However, high volume of data may not be processed by GovSchemAna app. Therefore to have an alternative arrangement for this, a subset of the full dataset may be chosen at random and applied to this app. With these abilities to perform analytics on a smartphone, it will definitely become more indispensable device than it was before.

References

[1] Alpaydin E., 2014, Introduction to machine learning, 3rd Edition, *MIT press*

[2] Chang C. C., Lin C. J., 2011, Libsvm: a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2, 3, pp.27.

[3] Collins C., Galpin M., Kappler M., 2011, Android in practice, *Manning Publications Co.*

[4] Das S. K., 2013, A brief scanning on performance of mahatma gandhi national rural employment guarantee act in assam, india, *American Journal of Rural Development*, 1, 3, pp.49-61

[5] Google Inc., 2017, Android, *URL https://www.android.com/*

[6] Google Inc., 2017, Google apis for android | google developers, *URL https://developers.google.com/android/*

[7] Hall M., 2017, Time series analysis and forecasting with weka, *URL http://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka*

[8] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H., 2009, The weka data mining software: an update, *ACM SIGKDD explorations newsletter*, 11, 1, pp.10-18.

[9] Haykin S., 2009, Neural networks and learning machines, 3rd Edition, *Pearson Education, Upper Saddle River, New Jersey*

[10] Meier R., 2012, Professional Android 4 application development, *John Wiley & Sons*

[11] Ministry of Electronics & Information Technology G. o. I., 2017, Open government data (ogd) platform india, *URL http://data.gov.in*

[12] Ministry of Electronics & Information Technology G. o. I., 2017, Schemes | national portal of india, *URL http://www.india.gov.in/my-government/schemes/*

[13] Ministry of Rural Development G. o. I., 2017, Mahatma gandhi national rural employment guarantee act, *URL http://www.nrega.nic.in*

[14] Mohanty S., Mishra A. K., Panda D. C., 2015, Igosa-a novel framework for analysis of and facilitating government schemes, *in: Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on, IEEE*, pp. 290-295.

[15] Mohanty S., Padhy S., 2017, A novel OFS-TLBO-SVR hybrid model for optimal budget allocation of government schemes to maximize GVA at factor cost, *Journal of Management Analytics*, pp.1-22.

[16] Pradhan A., Deshpande A. V., 2014, Composing Mobile Apps, Learn | Explore | Apply using Android, 1st Edition, *Wiley*

[17] Sahota A. S., 2008, Schemes on rural housing, *Rural Infrastructure: Sanitation, Housing, Health Care*, pp.102

[18] Satish S., Milne G., Laxman C., Lobo C., 2013, Poverty and social impact analysis of the National Rural Employment Guarantee Act in Karnataka to enable effective convergence, *World Bank, New Delhi*

[19] Saxena N., 2007, Rural poverty reduction through centrally sponsored schemes, *Indian Journal of Medical Research*, 126, 4, pp.381

[20] Sposaro F., Danielson J., Tyson G., 2010, iwander: An android application for dementia patients, *in: Engineering in Medicine and Biology Society (EMBC), 2010 annual international conference of the IEEE, IEEE*, pp.3875-3878

[21] University of Waikato, 2011, Weka-for-android, *URL https://github.com/rjmarsan/Weka-for-Android*

[22] Witten I. H., Frank E., Hall M. A., Pal C., 2016, Data Mining: Practical machine learning tools and techniques, 4th Edition, *Morgan Kaufmann*

A Brief Author Biography

Sabyasachi Mohanty – Sabyasachi Mohanty completed his Bachelor in Engineering from Institute of Technical Education and Research, Bhubaneswar, India in the year 2001. After working for 6 years as a software professional, he did his masters in Computer Science and Engineering from International Institute of Information Technology, Bhubaneswar, India in 2009. Till date, he is working as an academician with Centurion University of Technology and Management, Odisha, India, where he is also pursuing his Ph.D. His main areas of interest are information systems and machine learning.

Sudarsan Padhy – Sudarsan Padhy is currently Professor of Computer Science & Engineering at Silicon Institute of Technology, Bhubaneswar. Prior to this, he was Director at Institute of Mathematics and Applications, Bhubaneswar, India, and Professor of Mathematics at Utkal University, India. He obtained his Ph.D. degree in Mathematics from Utkal University in 1979 and pursued Postdoctoral research at University of Freiburg, Germany during 1980-81. He has over seventy published research papers and five books to his credit extending over fluid dynamics, finite difference and finite element method for solving partial differential equations, operation research, parallel algorithms, computational finance, machine learning, and computational biology.

M Vamsi Krishna – M Vamsi Krishna has completed M.Tech (CS) from Allahabad University, M.Tech (AI & R) degree from Andhra University, and received Ph.D. from Centurion University of Technology and Management, Odisha. Currently, he is working as HOD & Professor of Department of Computer Science and Engineering, Centurion University of Technology and Management. He has 17 years of experience in teaching & research. His research interests include Artificial Intelligence, Computer Networks, Image Processing, and Data Analytics.

