International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

Smart Expense Manager: Intelligent Personal Finance Management with Machine Learning

1st Aayush Patil
Dept. Information technology
School of Computing - MIT ADT UNIVERSITY
Pune, India
patilayush2610@gmail.com

2nd Ankit Lukadey
Dept. Information technology
School of Computing - MIT ADT UNIVERSITY
Pune, India
ankitrajput5618@gmail.com

3rd Hrishabh Kulkarni
Dept. Information technology
School of Computing - MIT ADT UNIVERSITY
Pune, India
hrikul01@gmail.com

4th Abhishek Samanta
Dept. Information technology
School of Computing - MIT ADT UNIVERSITY
Pune, India
abhisheksamanta09@gmail.com

5th Prof. Reetika Kerketta
Assistant Professor
Dept. Information technology
School of Computing - MIT ADT UNIVERSITY
Pune, India
reetika.kerketta@mituniversity.edu.in

ABSTRACT

This paper proposes a smart expense manager that incorporates machine learning (ML) techniques to analyze and track user money spending behavior. The system is designed to help users track their spending and identify areas where they can reduce expenses or increase savings. The system leverages ML algorithms to analyze user spending patterns and provide personalized recommendations based on their spending history. The proposed system also includes features such as budget tracking, expense categorization and transaction monitoring through the use of ML algorithms (Random Forest, Linear regression, boosting, neural networks), charts and graphical representation. The results of preliminary testing indicate that the proposed system is effective in helping users manage their expenses, with users reporting improved financial control and better financial outcomes. Overall, the proposed smart expense manager has the potential to improve financial literacy and promote more responsible financial behavior among users.

Keywords
Transactions; Machine learning; Random Forest; Linear regression; boosting; neural networks; financial behavior; expense manager.

1. INTRODUCTION

Managing daily spending is a major worry for everyone, and it's difficult to manage using manual processes, handwritten notes and paper receipts. We can improve the cost management process, decrease errors, and send refunds more quickly by employing software to track and monitor spending. The fundamental goal of tracking expenditure against income is to ensure that spending does not exceed available revenue. The first issue, as with tracking expenditure versus budget, is to determine whether the source of income has large surpluses or deficits. Hence, we proposed the solution, the smart expense manager. The smart expense manager allows users to view expenses on a monthly, weekly, and yearly basis. Visualization makes it easier to see the categories on which more money is spent. The expense management system also has the feature where you can set your monthly budget, so that users can spend money mindfully.

2. LITERATURE SURVEY

In modern times, machine learning is applied everywhere. The concept was derived from a few existing projects and worked on to analyze and develop this project so that it would potentially be of great assistance to users in managing their spending and investing wisely.

Z. Huber and H. Perrin created a project in which the system forecasts the following year's spending by analyzing data from the previous 2-3 months, which predicts the expenditure of each month of the upcoming year. [7]

W. Durand, M. J. Schmidt, R. Dupret, and P. Borreli created a system that forecasts future expenditures based on multiple expense categories.[8]

I. Savinkin and Mrs. M. Angel devised a system that assists users in reducing personal spending by categorizing expenses and determining the most expensive category. [9]

3. PROPOSED SYSTEM

In the proposed system we are trying to fetch messages from the user mobile, perform some pre-processing. After that the system will predict the expenditure of next month using one of the supervised machine learning algorithms.

FLOW DIAGRAM

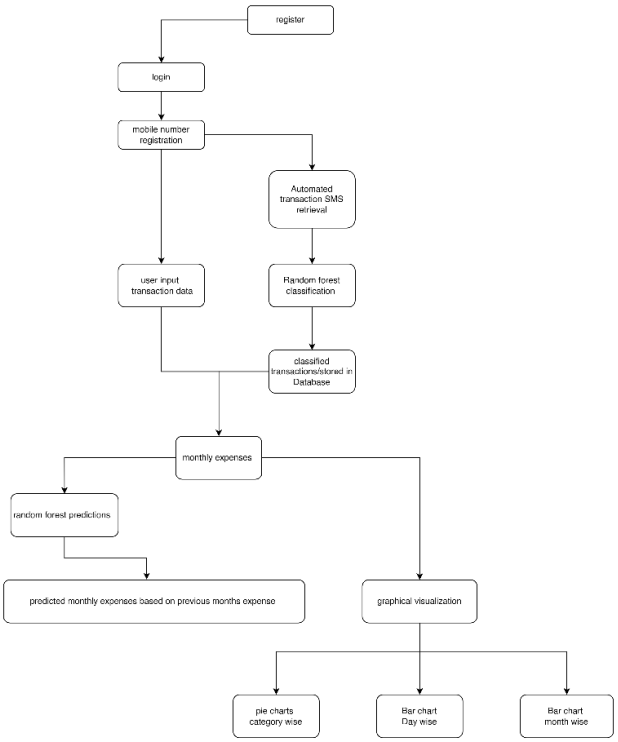

Figure 1. Application flow diagram

3.1 Data Extraction Methods

3.1.1 USER INPUT

An expense manager allows users to input data about their spending habits, including the date, category, amount, and possibly additional details such as the vendor or description of the purchase. Here are the steps that expense manager takes to process user input data:

1. Collect user input: The expense manager provides a user interface that allows users to input data about their expenses. in the form of text boxes, dropdown menus, or other interactive elements that allow users to enter information about their spending.

2. Validate input: Once the user has submitted their input, the expense manager validates the data to ensure that it is accurate and complete. This includes checking for missing or invalid fields, ensuring that the amount is in the correct format, and verifying that the date is valid.

3. Categorize expenses: The expense manager provides a set of predefined categories (e.g., groceries, entertainment, transportation) that users can assign to their expenses. This allows users to see how much they are spending in different areas of their life and identify areas where they may need to cut back.

4. Store data: Once the data has been validated and categorized, the expense manager stores it in a database. This allows users to view their spending history over time and generate reports on their spending patterns.

3.1.2 Automated SMS Retrieval

SMS retrieval: The expense manager app uses automated banking SMS retrieval technology to retrieve transaction information from the user's bank account via SMS. This can include the transaction amount, date, and vendor name. This data will be used for predictions.

Table 1. Dummy data set for user future spending amount and ml algorithm accuracy

User ID	Gender	Age	Income	Month	Spending Amount	ML Prediction
1	Male	25	40000	Jul	6000	6200
1	Male	25	40000	Aug	5000	4800
1	Male	25	40000	Sep	4500	4300
2	Female	35	60000	Jul	5500	5800
2	Female	35	60000	Aug	6000	6300
2	Female	35	60000	Sep	6500	6600
3	Male	45	80000	Jul	8000	7800
3	Male	45	80000	Aug	8500	8900
3	Male	45	80000	Sep	9000	9200
4	Female	30	50000	Jul	4000	4200
4	Female	30	50000	Aug	4500	4600
4	Female	30	50000	Sep	5000	5200
5	Male	40	70000	Jul	7000	7100
5	Male	40	70000	Aug	7500	7600
5	Male	40	70000	Sep	8000	8200
...	...	...	...	...	...	...

To establish how accurate a machine learning algorithm is at predicting future spending behaviors based on a given data set, the dataset must be divided into two parts: a training set and a testing set. The machine learning algorithm is trained using the training set, and its performance is evaluated using the testing set.
Several measures, including mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R-squared), can be used to assess the accuracy of the machine learning method. The specific metric utilized would be determined by the analysis's specific aims and constraints. It is important to remember that the accuracy of the machine learning algorithm is determined by several factors, including the algorithm used, the quality of the data, the features utilized in the model, and the size of the training set. To provide the best possible accuracy, these parameters must be thoroughly evaluated and optimized.

3.1.3 Studied Algorithms

1) Random Forest

Random Forest is a classifier that uses multiple decision trees on different subsets of a given dataset and averages them all to improve the dataset's predictive accuracy. The random forest takes the results from each decision tree and bases the final output on the majority votes of predictions rather than relying on a single tree.

2) Linear Regression
Regression is a technique for modelling a goal value using independent predictors. This method is commonly used for forecasting and determining cause and effect relationships between variables.

3) Gradient Boosting
Gradient boosting is a prominent and widely used machine learning approach for tabular datasets. It is capable of discovering any nonlinear link between your model's target and features.

4) Artificial Neural Network
ANNs are computer programs that can be taught to mimic correlations in the data sets. Once trained, the ANN can be used to predict the result on newly collected datasets.

All the above algorithms are studied and used to predict user's expenses based on previous expenses datasets to measure the accuracies of the following algorithms. To measure the accuracy of all these machine learning algorithms, several metrics can be used depending on the type of problem.

we used following metrics to evaluate the accuracies:

1. Random Forest:

• Regression problems: R-squared, mean squared error (MSE).

2. Linear Regression:

• Regression problems: R-squared, MSE.

3. Gradient Boosting:

• Regression problems: R-squared, MSE.

4. Artificial Neural Network (ANN):

• Regression problems: R-squared, MSE.

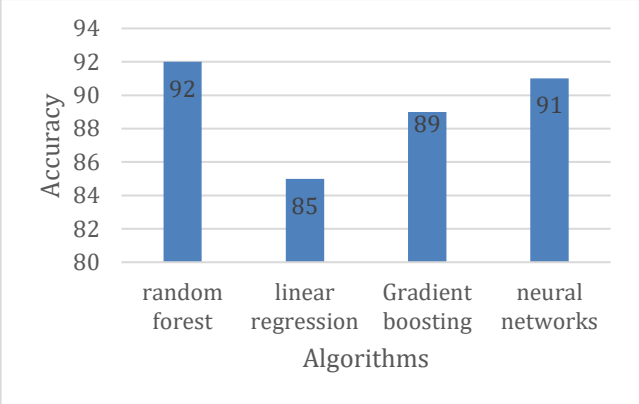

Figure 2. Relative Algorithm accuracy

```
In [50]: 1 model.intercept_
Out[50]: 34.21916368863002

In [56]: 1 from sklearn.metrics import mean_absolute_error , mean_absolute_percentage_error , mean_squared_error

In [57]: 1 mean_absolute_error(y_test,y_pred)
Out[57]: 3.1550309276024904

In [58]: 1 mean_absolute_percentage_error(y_test,y_pred)
Out[58]: 0.1635593588221794

In [59]: 1 mean_squared_error(y_test,y_pred)
Out[59]: 20.718012877838454
```

Figure 3. Random Forest accuracy findings

After studying and carefully evaluating all four algorithms we concluded that random forest provides the best fit results based on study conducted.

5. RESULTS

The derived conclusion from the performed research and experiment is that the **Random Forest** model can be accurate up to 70 to 80% in predicting future expenses, depending on the size and quality of data being used and can be used effectively to solve the proposed problem statement.

6. CONCLUSION

Referring to some of the research papers and implementing different machine learning algorithms on dummy dataset shown in Table. 1, we come to know that random forest algorithm will give best results compared to other machine learning algorithms. Refer figure 2.

7. ACKNOWLEDGEMENT

Professor Reetika Kerketta directed the research and development efforts. This project aided in the understanding of the various parameters involved in the analysis of a user's money spending behavior, as well as the machine learning techniques that can be used to predict a user's future expenses, in addition to the development of applications and the working and integration of front end and back end to create fully functional android and web applications.

8. REFERENCES

[1] V. Singh, L. Freeman, B. Lepri, and A. Pentland, "Predicting spending behavior using socio-mobile features".

[2] KIM Zi Won Comparison between different Reinforcement learning algorithms on open AI Gym environment (Cart-Pole v0) 2017.

[3] Ankit Choudhary A Hands-On Introduction to Deep Q- Learning using Open AI Gym in Python IIT Bombay EEE.

[4] Andrew NG Machine Learning Course Coursera. Available: https://www.coursera.org/learn/machine-learning

[5] J. Filliben et al., "Introduction to time series analysis", In NIST/SEMTECH Handbook of Statistical Methods, National Institute of Standards and Technology, 2003.

[6] P. Thanapal, Mohammed Yaseen Patel, T. P. Lokesh Raj and

J. Satheesh Kumar, "Income and Expense Tracker", Indian Journal of Science and Technology, Vol 8(S2), ISSN: 0974-5645(January 2014).

[7] Z. Huber and H. Perrin, "User expenditure prediction via machine-learning algorithms", Journal of Banking & Finance, vol. 34, no. 11, pp. 2767–2787, 2010

[8] W. Durand, M. J. Schmidt, R. Dupret, P. Borreli, "Category wise expenditure prediction model", Expert systems with applications, vol. 36, no. 4, pp. 7611–7616, 2009.

[9] I. Savinkin, Mrs. M.Angel, "monthly expenditure optimisation", Expert Systems with Applications, vol. 42, no. 4, pp. 1789–1796, 2015