# EXECUTIVE SUMMARY

**Deploying machine learning to understand saving habits of individuals within known varied financial measurement categories**

# Authored By

Rachita Patel National College of Ireland

### Research question

*How can unsupervised machine learning techniques help in discovering group of factors that influence the saving behaviour of the people in various financial measurement categories?*

* Data science methods like Multiple Linear Regression (MLR) and Random Forest (RF) to select attributes relevant to the six financial measurement categories.
* Then performing cluster analysis for those categories to find out which type of factors are responsible for a specific saving habit.

### Data used

To understand individuals’ saving habits, the financial well-being survey data from the Consumer Financial Protection Bureau (CFPB) has been used for this research. The goal of CFPB was to find the financial well-being score for the United States (U.S.) adults using this data. The CPFB financial well- being survey data was collected from the younger consumers (aged 18-61 years) and older consumers (aged 62 years and older) from the 50 U.S. States and Washington DC. It has 217 variables and 6394 records.

In addition to that, it collected details related to an individual’s cognitive psychology and financial planning attitude i.e. six categories of financial measurements. They are: (i) individual characteristics, (ii) household and family characteristics, (iii) income and employment characteristics, (iv) savings and safety nets, (v) financial experiences, and lastly (vi) financial behaviors, skills, and attitudes. Thus, the survey data related to these characteristics was useful for this research.

### Methods

The saving habit of individuals depends on various characteristics and the dataset chosen here comprises of attributes for all the categories i.e. 217 variables. So, the problems occurred when dealing with 217 attributes were: (a) clusters are not compact, (b) the number of clusters increases which causes misleading analysis, and (c) large processing time. Due to these reasons, the six categories of financial measurement were considered as reported in the CFPB financial survey report.

Thus, the entire data was divided into these categories and then evaluated. The outcomes of such a decision are: (a) better clusters, (b) saving habit for respective categories can be thoroughly analysed and understood, and (c) comparatively less processing time.

Now, each category has attributes relevant to that. But it is not necessary that all the attributes influence the saving habit of individuals. It is to be noted that this is a survey data, so it contains some variables which will not be relevant for this study. Thus, the influencing attributes are selected based on their: (a) correlation values, (b) p-values, (c) normalisation status, (d) outliers detection, and (e) using MLR and RF algorithms. The MLR and RF algorithms are evaluated using performance metrics like mean error (ME),

mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), and root mean square error (RMSE).

Those attributes which showed some relevance depending on the selection criteria and based on prior research were selected for cluster analysis. Clustering algorithms based on their simplicity and relevance were selected for this research. Algorithms like k-means, partitioning around medoids (PAM), clustering large applications (CLARA), hierarchical clustering, and fanny clustering were chosen for the analysis. The important step in clustering is to determine the optimal number of clusters and this cannot be selected randomly. So, the implementation for a varied number of clusters was conducted and then they were evaluated using internal validation metrics to check for their performance. The internal validation metrics includes connectivity, the Dunn’s index, and the Silhouette index. Based on that, the numbers of clusters relevant for a particular category were selected. Having known the number of clusters for a case, these results were thoroughly analysed to understand the type of factors responsible for a particular saving habit. Lastly, the saving habit and their influencers were interpreted for the respective financial measurement categories thereby understanding the saving behaviour.

### Conclusions

Depending on these results, inferences were drawn for the various factors leading to a saving habit. It was observed that income and education do have a positive impact on saving habits as noted by many researchers. Apart from that, married individuals have a less tendency to save than the unmarried ones. Similarly, individuals financially supporting children do not show good saving habit whereas retired individuals have the best savings. Also, house proprietors exhibit good saving behaviour. Apart from that, a good saving habit is found in the people who have not gone through any financial shock.

### Future Work

One of the major observations is that goal-oriented individuals have an increased tendency to save. It has been noted that there are individuals who save regularly are having $20k-$75k in their savings today and perhaps more can be saved. Primarily, this is possible if education is encouraged. Other than that, individuals with low income do not contribute much towards saving, so this group needs to be targeted to check for their disposable income and accordingly a saving schedule can be generated for them. Also, for individuals with dependents, saving habit has to be encouraged after knowing their profile. So, these varied types of behaviours or factors will further help in designing financial investment solutions for the people thereby contributing towards their financial wellness.

So, having access to the customer profile information and the actual banking transaction data will further help in analysing the individual and thereby a saving habit can be encouraged accordingly. By doing this, financial investment solution can be designed, financial wellness can be ensured, and there will be interaction between the bank and their clients. This can further lead to behavioral banking (FinTech era). Also, more than 50 percent of Irish people does show a good financial behaviour and this can be improved upon with the help of behavioral banking. This will have impact on the economic growth of the country thereby strengthening its financial stability.

**ABOUT THE AUTHOR**

**Rachita Patel, Msc in FinTech, B.E.**

## Having completed an undergraduate engineering in electronics and telecommunications from St. Francis Institute of Technology, Mumbai University, she then committed to develop her International career by relocating to Dublin. Rachita is now focused on deploying her knowledge of data science in financial services. Having contributed to projects in fields like machine learning, computer vision, robotics, security, and cryptography then she had a vision of enhancing financial sector with technology. She has a good knowledge of platforms like MATLAB, Python, and R.

*Contact details:*

(+353) 89 439 5449

[rachita.patel63@gmail.com](mailto:rachita.patel63@gmail.com) [linkedin.com/in/rachitapatel](lin)