Predictive Analytics for Startup Success among Women Entrepreneurs

Yuvarani D[1], Iyappan M[2]

[1]Assistant Professor, Department of Mathematics, Mahendra Arts & Science College (Autonomous), Namakkal, Tamilnadu, India
[2]Associate Professor, Department of Science, St. Francis College, Bengaluru-34, India

Abstract - **The study investigates factors influencing startup success among women entrepreneurs using predictive analytics techniques. A simulated dataset of 500 women entrepreneurs was generated, including variables such as age, education, location, industry, experience, initial capital, loans, grants, mentorship, and network size. Logistic regression, decision tree, and random forest models were applied to identify the most influential predictors of success. Results indicated that none of the predictors were statistically significant in the logistic regression, suggesting complex interactions between factors. Random forest importance scores highlighted initial capital, network size, and experience as key contributors to startup success. The findings provide insights for policymakers and support organizations to design tailored interventions to enhance women led startup success.**

Keywords - **Predictive Analytics, Women Entrepreneurs, Startup Success, Logistic Regression, Random Forest, Decision Tree**

I. INTRODUCTION

Women entrepreneurship has gained global attention due to its potential to drive economic growth and social development. Despite increased participation, women-led startups often face challenges such as limited access to capital, mentorship, and networks. Predictive analytics offers data-driven approaches to understand and forecast startup success by examining multiple factors simultaneously. This study aims to evaluate the impact of demographic, financial, and professional variables on the success of startups led by women using advanced statistical techniques including logistic regression, decision trees, and random forests.

II. RESEARCH DESIGN AND METHODOLOGY

Research Design

A quantitative, predictive research design was adopted using simulated survey data representing 500 women entrepreneurs across diverse industries and locations

Variables

TABLE I

Variable	Type	Description
Age	Continuous	Entrepreneur's age in years
Education	Categorical	Undergraduate, Graduate, Postgraduate
Location	Categorical	Urban, Semi-Urban, Rural
Industry	Categorical	Tech, Retail, Services, Manufacturing
Experience	Continuous	Years of entrepreneurial experience
InitialCapital	Continuous	Initial investment in INR
Loans	Binary	Received loan (1 = Yes, 0 = No)
Grants	Binary	Received grant (1 = Yes, 0 = No)
Mentorship	Binary	Received mentorship (1 = Yes, 0 = No)
NetworkSize	Continuous	Size of professional network
Success	Binary	Startup success (1 = Yes, 0 = No)

III. STATISTICAL ANALYSIS

Logistic Regression

A logistic regression model was fitted to identify significant predictors of startup success.

TABLE II

Predictor	Estimate	Std. Error	z value	p-value
(Intercept)	0.7548	0.7637	0.988	0.323
Age	-0.0122	0.01897	-0.641	0.522
Education Post graduate	-0.1589	0.2206	-0.72	0.471
Education Undergraduate	0.3439	0.2278	1.51	0.131
Location Semi-Urban	0.2842	0.2267	1.254	0.21
Location Urban	0.1182	0.224	0.528	0.598
Industry Retail	-0.1036	0.2615	-0.396	0.692
Industry Services	-0.3369	0.2533	-1.33	0.184
Industry Tech	-0.146	0.2558	-0.571	0.568
Experience	-0.0345	0.05365	-0.642	0.521
Initial Capital	-4.23E-07	7.01E-07	-0.603	0.546
Loans	0.0264	0.1902	0.139	0.89
Grants	0.1336	0.2003	0.667	0.505
Mentorship	-0.0425	0.1837	-0.231	0.817
Network Size	-0.0177	0.01825	-0.967	0.333

None of the predictors were statistically significant (p > 0.05), suggesting a complex interplay among factors influencing startup success.

Decision Tree Analysis

A decision tree was fitted using rpart to identify important splits for predicting success. Key splits were observed on Initial Capital, Network Size, and Experience.
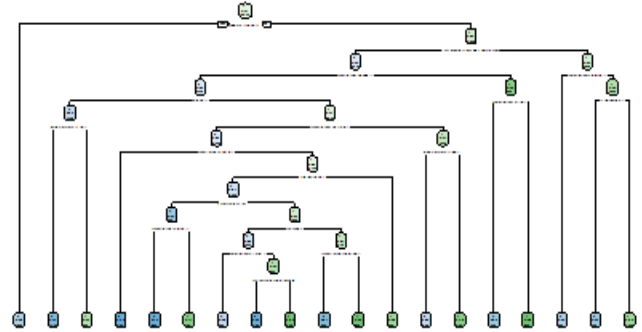

FIGURE I

Random Forest Analysis

A random forest model with 500 trees was fitted.
Variable importance (Mean Decrease Gini) indicated:

TABLE III

Predictor	Importance (Gini)
Initial Capital	58.96
Network Size	44.05
Age	42.03
Experience	28.84
Industry	18.71
Location	14.23
Education	13.91
Mentorship	8.8
Loans	8.29
Grants	7.86

Initial capital, network size, and experience are the most influential factors in predicting startup success.

IV. CONCLUSION

The predictive analytics approach highlighted that financial resources, professional networks, and entrepreneurial experience are critical for the success of women led startups.

Logistic regression alone may fail to capture complex interactions, while machine learning methods such as random forests provide more nuanced insights. These findings can guide policymakers, incubators, and mentors in designing targeted interventions to enhance startup success among women entrepreneurs.

REFERENCES

[1] Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. (2019). Multivariate Data Analysis (8th ed.). Cengage Learning.

[2] James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). An Introduction to Statistical Learning with Applications in R (2nd ed.). Springer.

[3] Kshetri, N. (2018). 1 Blockchain's roles in meeting key supply chain management objectives. International Journal of Information Management, 39, 80–89.

[4] OECD (2020). Women's Entrepreneurship: Policies and Practices. OECD Publishing, Paris.