Predictive Analytics in Professional Networking: Enhancing Connection Success Through Data-Driven Insights

Darwin Castellino
Engineering, Computer & Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
xrp0920@autuni.ac.nz

Mahsa Mohaghegh
Engineering, Computer & Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
mahsa.mohaghegh@aut.ac.nz

***Abstract*— This study leverages predictive modeling to enhance the success of professional connections within a networking platform, focusing on 188 significant user interactions. By analyzing features such as tenure differences, shared interests, and engagement metrics, and employing techniques like variance inflation factor (VIF) analysis and ANOVA, we refined our feature set and assessed predictor significance. To combat data imbalance, methods such as SMOTE and random sampling were utilized, ensuring robust model evaluations. Through comprehensive testing of models including Logistic Regression, Random Forest, Gradient Boosting, and SVM, the Random Forest model proved most effective, demonstrating superior precision and the ability to discriminate between successful and unsuccessful connections. Our findings illuminate the key drivers of networking success and underscore the potential of predictive analytics to significantly enhance user engagement.**

Keywords*— *Predictive Modeling, Professional Networking, Machine Learning, Random Forest, Social Network, Matching

I. INTRODUCTION

Networking is a pivotal professional skill that transcends basic social interactions, serving as a key driver for both personal career growth and organizational prosperity. It opens doors to the hidden job market, facilitating access to unadvertised opportunities through the development of meaningful relationships, which enable mentorship and career advancement [1]. Additionally, it fosters the exchange of unique ideas and expertise, enhancing individual skills and injecting novel perspectives into organizational challenges, thereby fostering an innovative and competitive culture[2]. Moreover, effective networking enhances professional visibility, leading to opportunities such as public speaking and collaborations that advance one's career [3]. Overall, networking involves creating and maintaining strategic relationships that benefit both individuals and organizations, proving essential for professional success and operational excellence. The landscape of networking has undergone significant evolution with the advent of digital technologies, transforming how professionals connect and interact. Platforms like LinkedIn and virtual meeting spaces have broadened the scope of networking, making it more accessible but also introducing challenges such as maintaining genuine connections in a predominantly digital space [4]. This digital shift has been accompanied by a diversification in the modalities of networking, encompassing traditional face-to-face interactions, online engagements, and hybrid formats. As organizations strive to maximize the benefits of both formal and informal networking, they increasingly adopt structured programs like mentorships and peer networking initiatives. These programs are tailored to harness the advantages of varied networking forms, adapting to the needs of a dynamic professional environment[3, 5]. This integration of different networking modalities helps organizations and individuals navigate the complexities of a digital age, ensuring that networking remains a robust tool for professional and organizational development.

Organizations increasingly leverage digital platforms to enhance networking through strategic initiatives. However, the effectiveness of such initiatives often hinges on the platform's ability to foster meaningful interactions that go beyond mere connectivity. By hosting diverse networking events and offering skill-enhancement training, these platforms provide varied interaction opportunities that bridge the gap between geographically dispersed teams.

Despite these efforts, optimizing these interactions to ensure successful professional connections remains a challenge. This highlights the need for advanced solutions to better predict and facilitate successful outcomes. This research directly addresses this challenge by exploring significant predictors of networking success used to match users effectively. Through the development and application of predictive models, this study aims to transform how professional networking platforms operate, ensuring they not only connect individuals but also foster substantial, rewarding professional relationships. This approach aims to significantly improve user interaction and connection success, leveraging data-driven insights to enrich the networking experience on these platforms.

In this research, we delve into the dynamics of professional networking within a platform used by a New Zealand-based company that orchestrates monthly coffee meets through random pairings. Utilizing a unique dataset that includes user interests, engagement metrics like question listings and contributions (rewarded with badges), organizational unit affiliations, message exchanges, and user join dates, we aim to uncover factors that contribute to the formation of meaningful professional connections in the absence of traditional demographic data. Our methodology integrates hypothesis testing and predictive modelling techniques to identify significant predictors of networking success. The anticipated impact of this research is substantial, enhancing the effectiveness of networking platforms by refining the underlying algorithms to foster more strategic and successful professional interactions. This advancement not only promises to deepen academic understanding of digital

networking dynamics but also offers practical strategies for organizations aiming to optimize the networking potential of their platforms.

The remainder of this paper is organized as follows: a comprehensive review of the literature on the factors influencing professional connections and the role of predictive analytics in networking, a detailed description of the methodology employed in this study, an analysis of the key findings, and a discussion of how these insights can be applied to improve the matching algorithms. Each section aims to build upon the last, culminating in a set of recommendations that could guide future enhancements to professional networking platforms.

II. Literature Review

A. Predictive Modeling in Social and Dating Platforms

There are some application of predictive models in online dating and social communication platforms which is similar to the scenario of matching users. [6] uses a Random Forest regression model to understand user satisfaction on Tinder, focusing on psychosocial traits. In contrast, [7] leverages CNNs for analysing user interactions on Tinder, enhancing message personalization and profile assessments. It extends this to a broader social context by evaluating male self-representation on Tinder through mixed methods, including qualitative interviews and quantitative experiments. These studies reveal the critical role of visual and textual content in user perception and interaction outcomes.

B. Advanced Recommendation Systems

There are researches on collaborative filtering in networking and social matching systems. [8] employs a Gaussian Mixture Model to update user profiles dynamically, optimizing social connections. [9] introduces EMUCF, a sophisticated collaborative filtering technique that uses non-linear similarity measures to enhance recommendation accuracy. Both articles emphasize the need for adaptive, responsive systems that reflect users' evolving preferences and behaviours.

C. Allocation and Matching Optimization

There are optimization techniques in different matching scenarios.[10] applies MILP and multi-attribute decision-making methods to roommate allocation in dormitories, aiming to enhance compatibility and reduce conflicts. [11] presents a hybrid model that combines collaborative and content filtering for mentor-mentee matching, integrating personality assessments to refine matches, illustrating the utility of hybrid models in enhancing personalization in both social and professional contexts.

D. Predictive Analytics in Education and Business

A number of researches showcase predictive modelling applications in educational and business settings.[12] utilizes an ANN to predict ERP system success in businesses, focusing on organizational characteristics. [13] uses machine learning to predict consumer behaviour on social media, providing insights for marketing strategies. [14] analyses factors influencing success in competitive surgical residencies, emphasizing geographical connections and experiential learning. [15] applies various machine learning models to predict startup success, highlighting the importance of industry characteristics.

In an overall sense, using predictive models seems the right approach towards our scenario, especially considering that traditional matching algorithms and recommendation systems might not fully address the unique complexities of professional networking. Predictive modelling allows for a more nuanced analysis of user interactions and behaviours, enabling the identification of subtle patterns that can predict successful connections with greater accuracy. This approach is particularly advantageous in the context of professional networking, where the quality of connections can significantly impact career development and professional growth.

Given the intricacies of professional relationships, which often depend on specific interests, industry backgrounds, and professional goals, conventional recommendation systems that rely purely on demographic data or superficial matching criteria might fall short. Predictive models, by contrast, can integrate a wide range of variables from user activity and communication patterns to derive sophisticated insights. These models can discern the likelihood of a fruitful professional interaction, making them invaluable for platforms aiming to enhance user engagement and networking success.

Therefore, in the context of enhancing professional networking platforms, predictive modeling emerges as a superior strategy that can transform passive networking environments into active, engaging communities that foster meaningful professional relationships. This shift towards data-driven, predictive approaches is essential for developing networking solutions that are both effective and scalable, ensuring users gain the most value from their interactions on the platform.

III. Methodology

A. Data Collection and Preprocessing

The data collection focused on 188 unique professional connections from a company's data management system, where direct messages were exchanged.

This dataset was compiled from several CSV files: the first file contained detailed connection records, including user matchings and match dates. A second file provided comprehensive user profiles, listing interests, joining dates, mentor status, and engagement metrics such as contributions and listings, which contributed to users' badge levels. Messages exchanged between users were extracted from a third file, with the inclusion criterion based on the timing of the messages—only those sent between the current and the next draw dates were considered relevant to a particular connection.

These disparate data sources were meticulously linked using unique user IDs to create a master file that included complete user details and their communication records for each connection. Importantly, due to compliance with privacy regulations, no demographic information such as age, gender, or job titles was included, ensuring the study adhered to strict privacy standards. The final dataset, thus, comprised a unique connection ID for each pair, enriched with both users' detailed information and their message exchanges, providing a solid foundation for analysing the dynamics influencing successful professional interactions on the platform.

This dataset forms the foundation for the next phase, focusing on feature engineering to create new columns by analyzing the interactions between user attributes.

TABLE I. FEATURE ENGINEERING SUMMARY

Sr.no	Column Created	Calculation	Description
1.	Shared Interests	Sum of matching interests between both users	Indicates the total number of shared interests, suggesting compatibility.
2.	Tenure Difference Days	Absolute difference in days between join dates	Measures the experience gap between the users, potentially affecting interaction dynamics.
3.	Badge Score Ratio	Badge scores of User1 / Badge scores of User2	Represents the relative engagement level of the users on the platform.
4.	Mentor Ratio	Average of mentor status (0 or 1) for both users	Reflects whether one or both users are mentors, indicating influence and leadership.
5.	Profile Completion Ratio	Average of profile completion scores for both users	A measure of how completely both users have filled out their profiles.
6.	Same Organization Unit	1 if both users are from the same organization unit, 0 otherwise	Identifies if the users are from the same organizational unit, which can influence their networking ease and effectiveness.
7.	Average Contribution Badge	Average contribution badges of both users	Quantifies the average contributions made by the users, a proxy for user activity level.
8.	Average Listing Badge	Average listing badges of both users	Averages the recognition each user has received for their listings, indicative of their visibility on the platform.
9.	Average Number Of Connections	Average number of past successful connections for both users	Provides insight into the users' networking history.

B. Feature Engineering

In feature engineering, new variables were derived from the existing data to better capture the dynamics and potential predictors of successful professional connections. This process involved combining various attributes from both users involved in each connection to create meaningful indicators that could illuminate the underlying factors influencing networking outcomes. Below is a detailed table 1 outlines the columns created during this phase:

This table 1 summarizes the calculated features engineered from the raw data, each designed to provide insights into various aspects of user behavior and their interactions within the professional networking platform. These new variables are instrumental in developing a deeper understanding of what drives successful connections, forming the basis for subsequent predictive modeling efforts.

C. Text Mining for Connection Success

Text mining was employed in this study to analyze the exchanges between users on a professional networking platform, aiming to identify whether these interactions led to successful outcomes such as meetings or significant engagements. The process began with aggregating message data from exchanges linked to specific connection, consolidating all messages for each connection into a single document to maintain the context and continuity of conversations. Using the zero-shot classification technique, from Hugging Face's Transformers library, each aggregated text was categorized into "meeting set" or "meeting not set." This method was chosen for its ability to classify texts accurately without needing task-specific training data, ideal for handling the diverse and dynamic content of user messages.

The analysis revealed that 89.4% of connections were classified as successful, where a meeting or substantial interaction was arranged, while 10.6% were not, as visually depicted in the accompanying table 2.

TABLE II. OUTCOMES FROM TEXT MINING ANALYSIS

Outcome	Count	Percentage (%)
Success	168	89.4
No Success	20	10.6

D. Correlation Analysis

The Pearson correlation coefficient revealed key relationships such as a high positive correlation between Badge score of users 1 and Average Listing Badge ($r = 0.753$), indicating that users with higher engagement scores tend to have more visibility through listings. Additionally, Tenure difference days and Badge score ratio exhibited a moderate positive correlation ($r = 0.297$), suggesting that differences in platform experience between users correlate with disparities in engagement levels.

E. Variance Inflation Factor (VIF) Analysis

As shown in table 3 initial VIF values indicated high multicollinearity, particularly for Badge Scores and Average Contribution Badge, both showing infinite values, suggesting perfect multicollinearity. This led to the decision to drop these features to enhance model stability and reliability. The revised VIF values post-adjustment showed acceptable levels for all included features, significantly reducing multicollinearity concerns.

F. ANOVA Testing

One-way ANOVA was employed to test for differences in means between successful and unsuccessful connections across various features. Tenure Difference was the only feature that showed a statistically significant difference (F-statistic: 5.0738, p-value: 0.0255), suggesting it significantly influences connection success. Other features did not exhibit statistically significant differences, which could imply that factors such as shared interests and engagement levels might not directly impact the success of connections in this context.

TABLE III. VIF ANALYSIS

Feature	VIF Before	VIF After
Shared_Interests	3.465211	3.285886
Tenure_Difference_Days	3.122679	3.084614
Badge_Score_1	inf	-
Badge_Score_2	inf	-
Badge_Score_Ratio	19.967134	3.897483
Mentor_Ratio	2.375110	2.315469
Profile_Completion_Ratio	25.070746	-
Same_Org	1.260530	1.227175
Average_Cont_Badge	inf	-
Average_List_Badge	inf	6.115330
Average_No_Of_CC	11.506106	5.439751

G. Predictive Modeling and Evaluation

This phase of the research focuses on developing predictive models to assess and predict the success of professional connections within a networking platform. Utilizing insights derived from our feature engineering and correlation analysis, we aim to employ several machine learning algorithms to understand and predict successful networking outcomes. This approach builds on existing literature that explores predictive modeling in social and dating platforms, recommendation systems, and various professional contexts. By adapting methodologies similar to those used in studies on online platforms, our goal is to refine predictive capabilities to enhance user engagement and satisfaction within professional networks.

- Data Balancing

As revealed in Table 2, the data showed a significant imbalance with 89.4% of connections being successful and only 10.6% unsuccessful. This imbalance can lead to biased predictive performance, where models might overly favor the majority class. To address this, data balancing is crucial as it ensures that both classes are adequately represented, enhancing the model's ability to generalize across different scenarios. The Synthetic Minority Oversampling Technique (SMOTE) was employed to balance the dataset by generating synthetic samples for the underrepresented class. This method was chosen for its effectiveness in creating realistic samples, thus providing a balanced basis for training robust models.

- Model Selection and Evaluation

Reflecting on the literature review, models similar to those used in enhancing social and dating platforms were selected. Models such as Random Forest, Gradient Boosting, and Support Vector Machines were chosen due to their proven capabilities in handling complex datasets with imbalanced classes. Random Forest, for instance, is noted for its high accuracy and robustness against overfitting, making it suitable for this dataset with diverse features. Gradient Boosting was selected for its strength in sequential correction of predecessors' errors, optimizing for precision and minimizing false negatives, a key aspect in predictive outcomes. Support Vector Machines were included for their effectiveness in high-dimensional spaces, crucial for datasets with numerous engineered features.

The insights gained from this modeling will directly contribute to enhancing the matching algorithms used on the platform, ensuring more tailored and effective user connections. By integrating the predictive insights into the platform's operational framework, we aim to improve the quality of matches, thereby increasing user satisfaction and engagement.

The dataset was then split into a training set (70%) and a testing set (30%). Each model was trained using the balanced training set, where the algorithms adjusted their parameters to best fit the data. The models learned to identify patterns and make inferences based on the training data, which included engineered features like the differences in tenure, the level of user engagement represented by badges, and the similarity in professional interests. Once trained, the models were tested using the separate testing set, which evaluated their performance based on several metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. These metrics were critical in understanding each model's strengths and weaknesses in predicting connection success.

TABLE IV. PREDICTIVE MODEL PERFORMANCE COMPARISON

Model	Performance Parameters				
	Accuracy	***Precision***	***Recall***	***F1 Score***	***AUC-ROC***
Logistic Regression	0.6237	0.6800	0.6071	0.6415	0.7250
Random Forest	0.9108	0.9795	0.8571	0.9142	0.9801
Gradient Boosting	0.8415	0.8333	0.8928	0.8620	0.9174
Support Vector Machine	0.6336	0.7317	0.5357	0.6185	0.6440

As shown in table 4, Random Forest model excelled, showing high scores across most metrics, indicating its effectiveness in handling the dataset's complexities. Similarly, Gradient Boosting showed strong performance, especially in precision and F1 score, affirming its capability to reduce false positives and enhance prediction accuracy.

IV. RESULTS AND DISCUSSION

This research delves into various analytical methodologies to unearth insights that could significantly enhance networking strategies in real-world applications. By understanding the factors that influence networking success, platforms can tailor their functionalities to foster more meaningful professional interactions.

The text mining analysis revealed that a majority (89.4%) of the message exchanges between users culminated in successful connections, suggesting that active communication is a strong predictor of positive networking outcomes. This observation underscores the importance of engagement and interaction within networking platforms. In practice, this could guide platform developers to design features that encourage more frequent and substantive communication among users, such as prompting users to initiate conversations or suggesting topics of mutual interest to both parties.

In the correlation analysis, it was found that users who are involved in mentoring tend to be more visible on the platform, and those differences in how long users have been members are associated with how actively they engage. These moderate correlations, while not strong, are significant enough to suggest that mentoring could boost user participation and that customizing user experiences based on their membership duration could enhance engagement. This information could help networking platforms develop strategies that encourage mentoring and tailor experiences to meet the needs of both

new and seasoned users, thereby improving overall community dynamics and user satisfaction.

The ANOVA results pointed out that the difference in tenure between users (Tenure_Difference_Days) has a statistically significant impact on the success of networking connections. This insight is particularly crucial as it highlights how the length of time users have been active on the platform can influence their networking experience. Users with longer tenure may have more extensive networks and greater familiarity with the platform's functionalities, which can aid them in forging successful connections. Platforms can leverage this information by pairing newer users with more experienced ones in mentorship programs or creating newcomer-friendly networking events that aim to quickly integrate them into the community.

Among the predictive models tested, the Random Forest model stood out with its exceptional performance, particularly in terms of precision and recall. Its ability to handle complex and nonlinear relationships between features makes it especially suited for predicting successful networking outcomes where multiple factors are at play. This model could be integrated into networking platforms to predict the likelihood of success between potential connections, helping to match users whose interactions are most likely to result in successful outcomes based on their engagement patterns and profile similarities.

The application of such a predictive model in a real-world networking platform could revolutionize how connections are suggested. By analysing past interaction data and user profiles, the model can identify patterns and characteristics of successful interactions, which can then be used to make data-driven suggestions for future connections. This would not only enhance the user experience by increasing the chances of successful networking but also make the platform more efficient by effectively utilizing data to foster genuine professional relationships.

V. Conclusion

This research has provided valuable insights into the factors influencing successful professional networking by utilizing text mining, correlation analysis, and predictive modelling. The text mining analysis revealed that active communication significantly correlates with networking success, suggesting that platforms should encourage more meaningful interactions among users. Correlation analysis highlighted that mentoring involvement and longer platform tenure are linked to higher user engagement. These findings can help tailor user experiences, promoting mentoring opportunities and adjusting features based on user tenure to boost engagement and success rates. Among the predictive models evaluated, the Random Forest model was notably effective, underscoring its potential for enhancing user matching processes on networking platforms. Integrating such models could refine how connections are suggested, potentially increasing successful user interactions and overall satisfaction.

Future research could explore longitudinal studies to assess how user behaviours evolve and impact networking success over time. Additionally, expanding the research to include a more diverse range of demographics could improve the understanding of universal and niche factors affecting networking efficacy. Implementing adaptive machine learning models that evolve based on user interaction data could also offer more personalized and responsive networking experiences.

Testing these models and strategies in real-world scenarios would validate their effectiveness, providing a practical blueprint for enhancing professional networking platforms. This continued exploration will help develop more sophisticated tools that cater to the dynamic needs of users, fostering richer professional relationships and community engagement within the platform.

References

1 De Klerk, S.: 'The importance of networking as a management skill', South African journal of business management, 2010, 41, (1), pp. 37-49

2 Porter, C.M., Woo, S.E., Alonso, N., and Snyder, G.: 'Why do people network? Professional networking motives and their implications for networking behaviors and career success', Journal of Vocational Behavior, 2023, 142, pp. 103856

3 Gibson, C., H. Hardy III, J., and Ronald Buckley, M.: 'Understanding the role of networking in organizations', Career Development International, 2014, 19, (2), pp. 146-161

4 Bazine, N., Serra, J., Giunchi, M., and Peña-Jimenez, M.: 'The Negative Consequences of Networking through Social Network Services: A Social Comparison Perspective', Computers in Human Behavior, 2024, pp. 108456

5 Chakraborty, P.: 'It'sa match! Or is it? A lab experiment on mentorship', in Editor (Ed.)^(Eds.): 'Book It'sa match! Or is it? A lab experiment on mentorship' (Mimeo, 2019, edn.), pp.

6 Vera Cruz, G., Aboujaoude, E., Rochat, L., Bianchi-Demichelli, F., and Khazaal, Y.: 'Finding intimacy online: A machine learning analysis of predictors of success', Cyberpsychology, Behavior, and Social Networking, 2023, 26, (8), pp. 604-612

7 Bloshenkina, A.A., and Tcyguleva, K.V.: 'Application of Machine Learning to Create a Recommendation in Social Communication Based on Data Analysis', in Editor (Ed.)^(Eds.): 'Book Application of Machine Learning to Create a Recommendation in Social Communication Based on Data Analysis' (IEEE, 2021, edn.), pp. 241-245

8 Dana, I., Agarwal, U., Ajay, A., Rastogi, S., and Alkhayyat, A.: 'Recommendation Mechanism to Forge Connections Between Users with Similar Interests', in Editor (Ed.)^(Eds.): 'Book Recommendation Mechanism to Forge Connections Between Users with Similar Interests' (Springer, 2023, edn.), pp. 455-470

9 Jain, A., Nagar, S., Singh, P.K., and Dhar, J.: 'EMUCF: Enhanced multistage user-based collaborative filtering through non-linear similarity for recommendation systems', Expert Systems with Applications, 2020, 161, pp. 113724

10 Khalili-Fard, A., Tavakkoli-Moghaddam, R., Abdali, N., Alipour-Vaezi, M., and Bozorgi-Amiri, A.: 'A roommate problem and room allocation in dormitories using mathematical modeling and multi-attribute decision-making techniques', Journal of Modelling in Management, 2024

11 Varghese, J., and Mohaghegh, M.: 'Personality-Based Hybrid Machine Learning Model for Mentor-Mentee Matching using Collaborative and Content Filtering Methods', in Editor (Ed.)^(Eds.): 'Book Personality-Based Hybrid Machine Learning Model for Mentor-Mentee Matching using Collaborative and Content Filtering Methods' (IEEE, 2023, edn.), pp. 1-6

12 Rouhani, S., and Ravasan, A.Z.: 'ERP success prediction: An artificial neural network approach', Scientia Iranica, 2013, 20, (3), pp. 992-1001

13 Chaudhary, K., Alam, M., Al-Rakhami, M.S., and Gumaei, A.: 'Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics', Journal of Big Data, 2021, 8, (1), pp. 73

14 Nasser, J.S., Artino Jr, A.R., Kind, T., Duan, X., Mihalic, A.P., and Chretien, K.: 'Matching into competitive surgical residencies: predictors of success', Medical education online, 2023, 28, (1), pp. 2189558

15 Kim, J., Kim, H., and Geum, Y.: 'How to succeed in the market? predicting startup success using a machine learning approach', Technological Forecasting and Social Change, 2023, 193, pp. 122614