Detecting Fake Profiles on Online Matrimony

Student Name: Vaibhav Garg

IIIT-D-MTech-CS-GEN-MT17064

May, 2019

Indraprastha Institute of Information Technology
New Delhi

Thesis Committee
Dr. Ponnurangam Kumaraguru (Chair)
Dr Arun Balaji Buduru
Dr. Siddhartha Asthana

Submitted in partial fulfillment of the requirements
for the Degree of M.Tech. in Computer Science,
in General Category

Keywords: Spam, Fake Profile, Online Matrimony

Certificate

This is to certify that the thesis titled **"Detecting Fake Profiles on Online Matrimony"** submitted by **Vaibhav Garg** for the partial fulfillment of the requirements for the degree of *Master of Technology* in *Computer Science & Engineering* is a record of the bonafide work carried out by him under our guidance and supervision in the Security and Privacy group at Indraprastha Institute of Information Technology, Delhi. This work has not been submitted anywhere else for the reward of any other degree.

Dr. Ponnurangam Kumarguru

Indraprastha Institute of Information Technology, New Delhi

Abstract

In a diverse country like India, socio-economic factors like religion, caste, language, income along with other common physical, professional based factors, play a vital role while searching for a spouse. With the surge of Internet connectivity, online matrimonial websites have become hugely popular to cater such needs. Most of the users registered on these portals have genuine intention of finding their desired life partner, however due to various factors, it attracts few profiles with no genuine intention for the same. Such profiles are also known as fake profiles. These profiles lead to bad user experience as well as revenue loss for the online matrimony business. To dig into this problem, we have chosen a use case of India's leading matrimony site and studied the behaviour, edit and profile differences between fake and genuine accounts. In this thesis, we present a machine learning based approach to identify such fake profiles on online matrimony. Due to lack of labelled examples for in-genuine users, we solve the above problem as anomaly detection problem. In this thesis, we use autoencoder which is widely used algorithm for anomaly detection. We capture user's behaviour, profile information and edit history to predict him/her as in-genuine or genuine profile. We then treat this problem as a reconstruction task using autoencoder which is trained on a set of genuine profiles features. While prediction, the autoencoder shows small reconstruction error for genuine profiles and a very high reconstruction error for the fake profiles and detect them. The proposed system produces 91.76% accuracy with 90.2% recall for fake class. To the best of our knowledge, this is the first study done to detect fake profiles in online matrimony domain.

Acknowledgments

It is my privilege to express my sincerest gratitude to my advisor, Dr. Ponnurangam Kumaraguru for giving me this opportunity to work on this thesis. I would also like to thank him for his valuable inputs, guidance, encouragement and wholehearted support throughout the thesis. I would like to thank my esteemed committee members, Dr Arun Balaji Buduru and Dr Siddhartha Asthana for agreeing to evaluate my thesis work. This work is in collaboration with InfoEdge India Ltd. I would like to thank Adhish and Hunny who were the driving force of this project. Special thanks to Aman Sharma from InfoEdge who has helped me in visualization of data. I am also grateful to all the members of my Precog family at IIIT Delhi who have consistently helped me with their inputs and suggestions on the work. Last but not the least, I would like to thank all my supportive family and friends who encouraged me and kept me motivated throughout the thesis.

Contents

List of Figures

Chapter 1

Introduction

1.1 About Online Matrimony

India is a land of diverse religions, castes, cultures and languages. While finding one's life partner, most of the people give very high importance to socio-economic factors like income, religion, mother tongue along with physical and professional attributes like age, height, weight, education, etc. According to the Indian Human Development Survey [27] conducted on Indian population living in urban cities, only 5% of respondents have done inter-caste marriage. The same study found out that inter-religion marriages were even less common in urban India. As per the Lok Foundation-Oxford University surveys conducted on Indian households [26], three-quarters of respondents said that they would not accept an inter-caste marriage for any of their children. This validates the fact that in India, factors like caste and religion play important role while choosing one's life partner.

Online matrimony is the web platform which helps the user to find his/her desired life partner based on these factors. On these platforms, users can look for the people (who are matching his/her desired set of attributes) and facilitate them to send interest message to each other's profile in order to initiate conversation.

Online matrimony helps its users to search for their potential spouse via different filters. These filters are specified on attributes like mother tongue, caste, religion, occupation, education, etc. A user gets the recommendation of other users based on filters specified by him/her. Moreover, online matrimony caters the need of different categories of people (like non-married and divorcees) and provides them an unbiased interface where they can easily search and contact for any particular category of people.

1.2 Functionalities Provided by Online Matrimony

Following are the user functionalities provided by online matrimony:

- A user can register himself on the platform by specifying his own socio-economic and physical attributes like mother tongue, marital status, city, income, age, etc.
- A user can also specify the desired profile attributes which he/she would like to have in his/her life partner. Based on desired profile attributes (DPA) and user's own attributes, he/she gets recommendations of his/her potential spouse.
- After registration, the user can send interest messages to others as a token of liking other's profile. The receiver of interest can then reply in affirmation or refutation.
- Based on mutual acceptance, a paid user can initiate a conversation with a profile of his/her interest.
- A user profile is able to edit his/her profile attributes like age, income, education to update their details and receive similar recommendations.

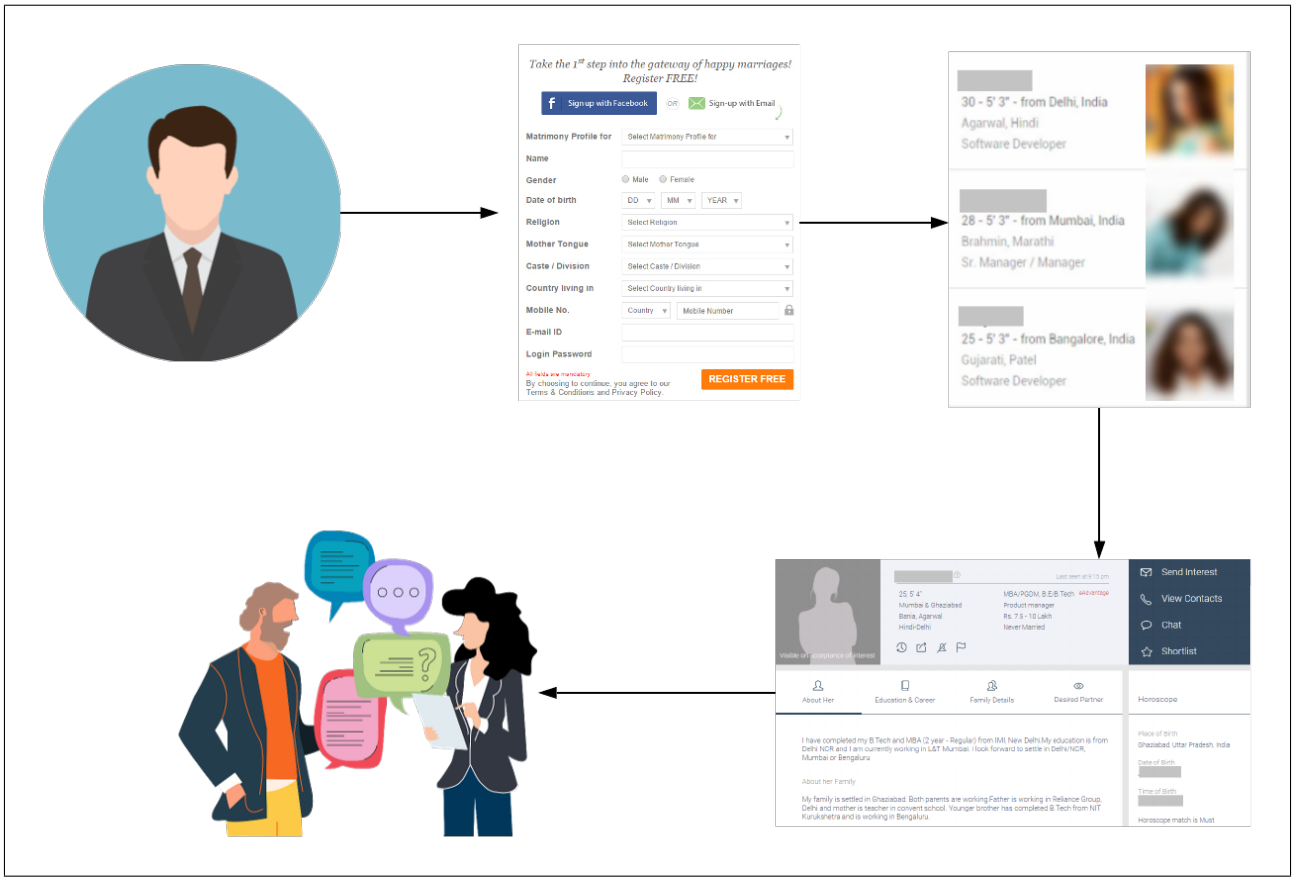

Figure 1.1: User first registers himself on the matrimony portal. He then gets profile recommendations and can view those profiles based on his interest. If he likes someone's profile, then he can send her interest message and initiate conversation.

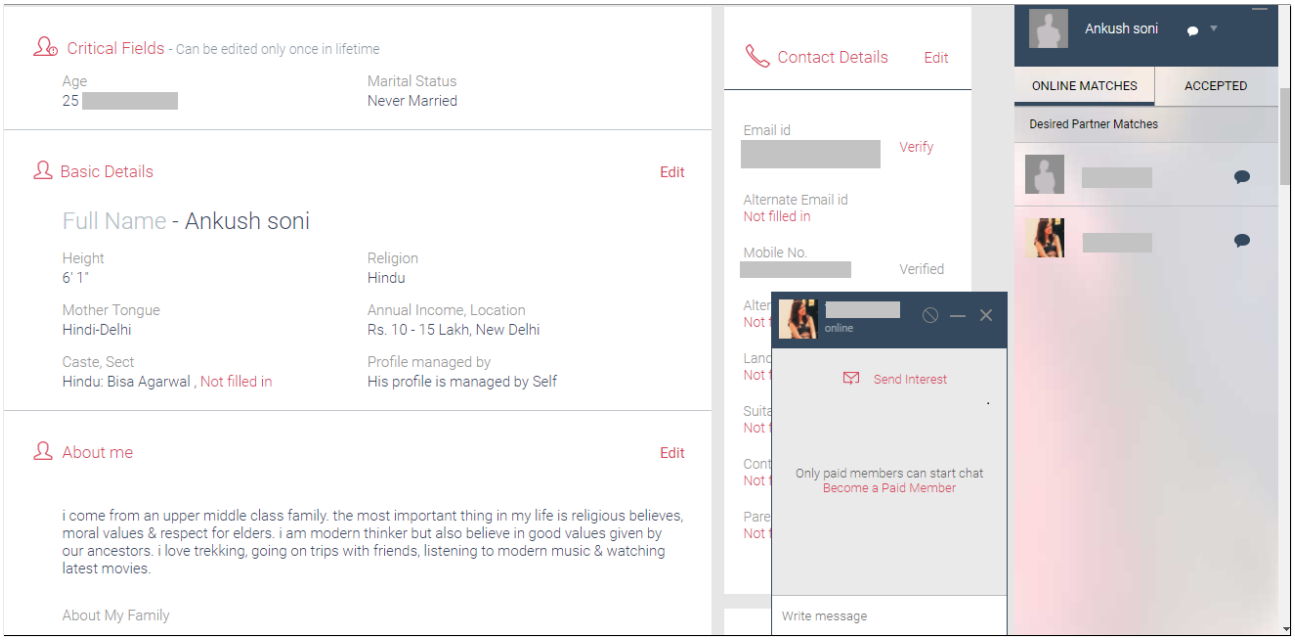

Figure 1.2: User can specify his socio-economic, physical and professional attributes while creating his profile on online matrimony. Matrimony platform also facilitates its users to start conversation with peer users who fall in his interests.

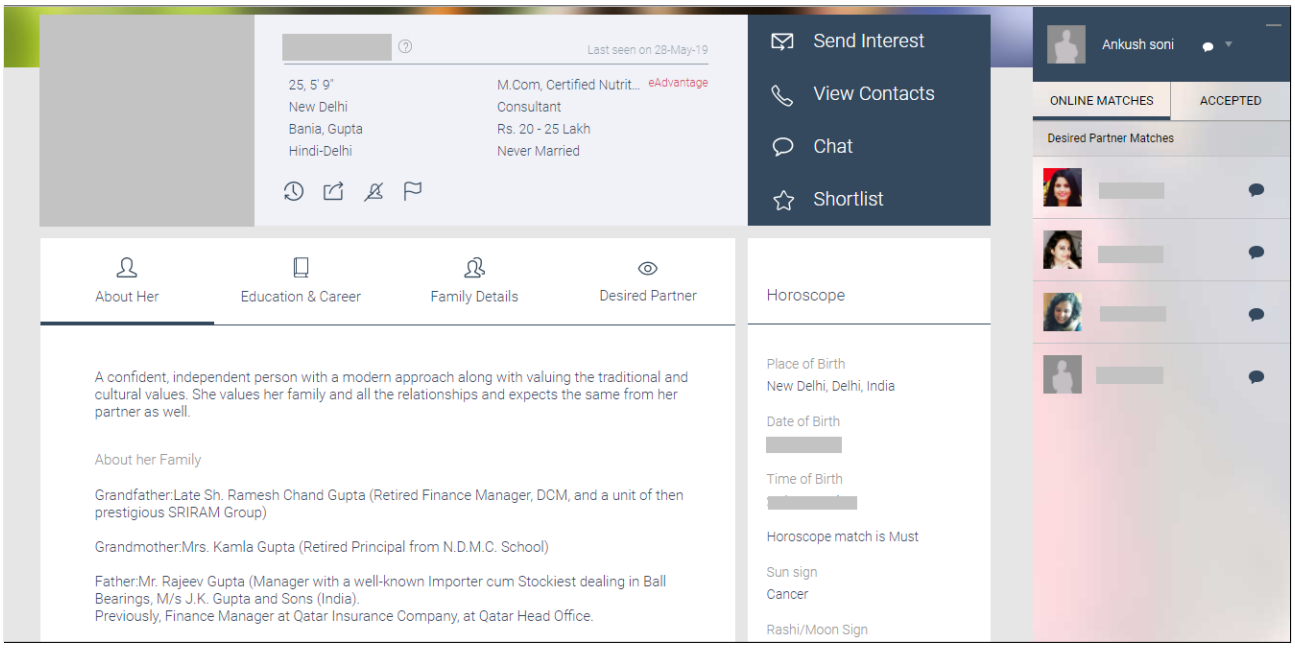

Figure 1.3: Based on the recommendations received by user on the portal, he can view the profiles which he finds suitable for marriage and send interests to them.

User's life cycle on a popular Indian matrimony portal is shown in Figure 1.1. User upon successful registration on the portal, gets recommendations for his/her perspective partner. He/She can view the full profile details of the recommended members and send them interest messages. If the user finds their profile suitable for marriage, then can gradually start conversation with them. Figure 1.2 shows user's own profile on online matrimony. We can see that a user has to specify his/her socio-economic attributes like religion, mother tongue income and physical attributes like age, height etc to create a profile on the portal. User can also edit these specified profile attributes at later point in time. Moreover, Figure 1.3 shows how a prospective partner's profile looks on matrimony portal. User can see partner's attributes and have an option to send interest to him/her. Portal also facilitates the user to shortlist different users based on his/her criteria and later select one out of them for marriage.

Chapter 2

Problem Statement and Motivation

2.1 What are fake profiles and how they affect the platform

On a matrimonial portal, most of the users have a genuine intention of finding his/her desired life partner. However, due to various factors like monetary benefits and data crawling, it attracts few people with malicious intentions. Some profiles have been encountered on online matrimony which try to contact many people and gradually ask them for money or other benefits [19]. This act shatters the other user both emotionally and monetarily and as a result, users start losing trust from the site. Moreover, other marriage bureaus (in competition) also create such profiles on the platform to contact several people and insist them to subscribe to their organization. Thus the existence of such fake profiles on the platform is highly undesirable to the business and other genuine users.

Figure 2.1 shows one such fake profile which exhibits high inconsistency between attributes like religion (Hindu: Brahmin) and mother tongue (Urdu). Moreover, in this case education (High School), occupation (IT Engineer) and income (100,001 US dollars) do not comply with each other. A similar trend has been shown in figure 2.2 for another fake user.

Figure 2.3 and 2.4 show the abnormal edit pattern exhibited by a fake user along education attribute. Initially on the portal, a fake user specifies his education qualifications as B.Tech and M.Tech. However, after few days, he changes them to a Charted accountant. Apart from this, there are also cases when a fake user shows behaviour heterogeneity while sending interests on the portal. Generally, these fake profiles are detected by manually scrutinizing all active users, which consumes substantial many man hours and resources. Thus there is need of an hour to build a machine learning based system which can identify fake users present on the portal.

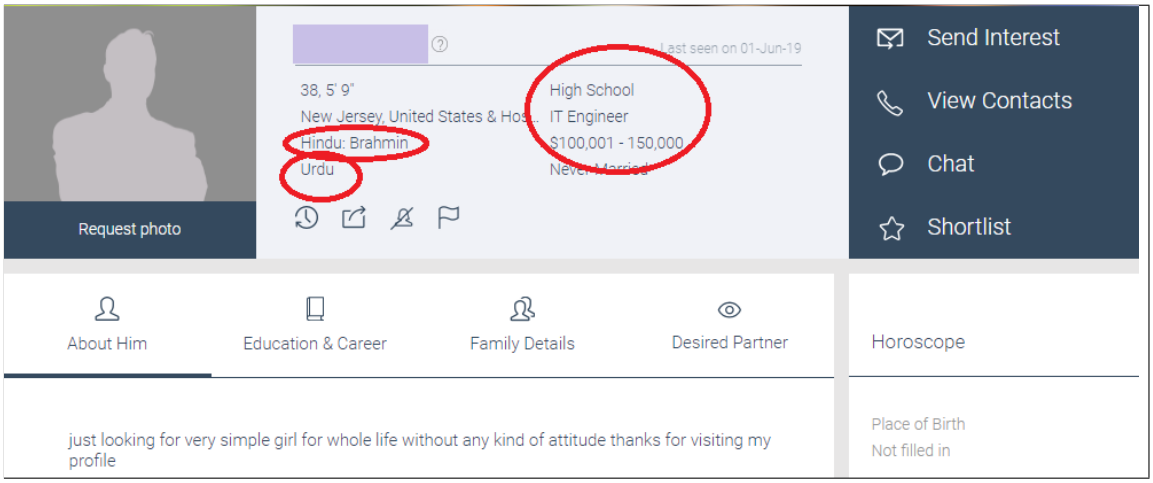

Figure 2.1: Fake profile showing high profile inconsistency along religion, mother tongue, education and income attributes. It is unlikely that such combination of categorical attributes co-exist.

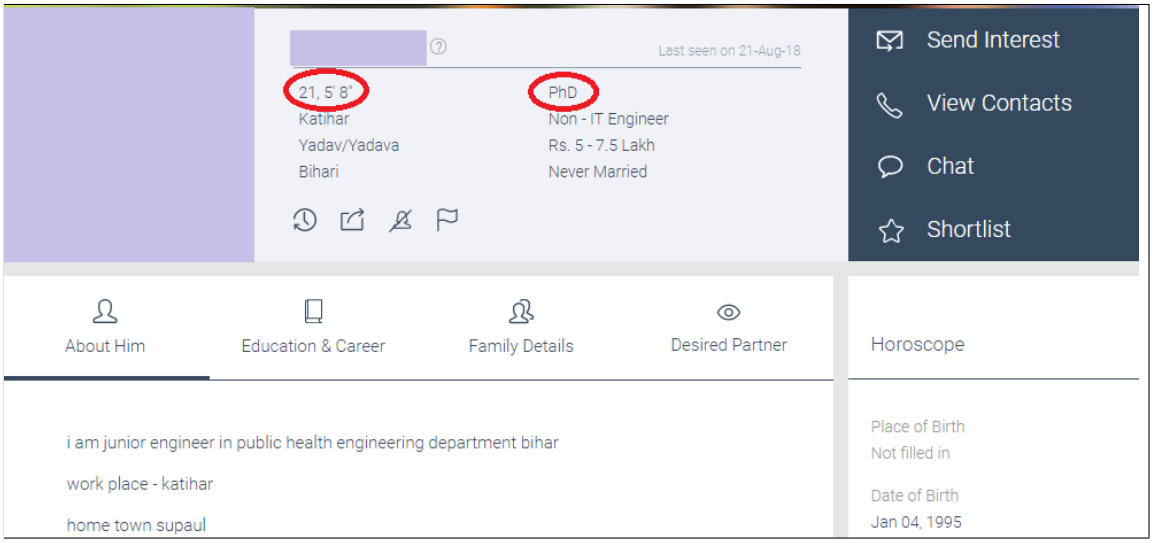

Figure 2.2: Fake profile showing profile inconsistency along age and education attributes. It is unlikely that a person of 21 years old hold a PhD degree.

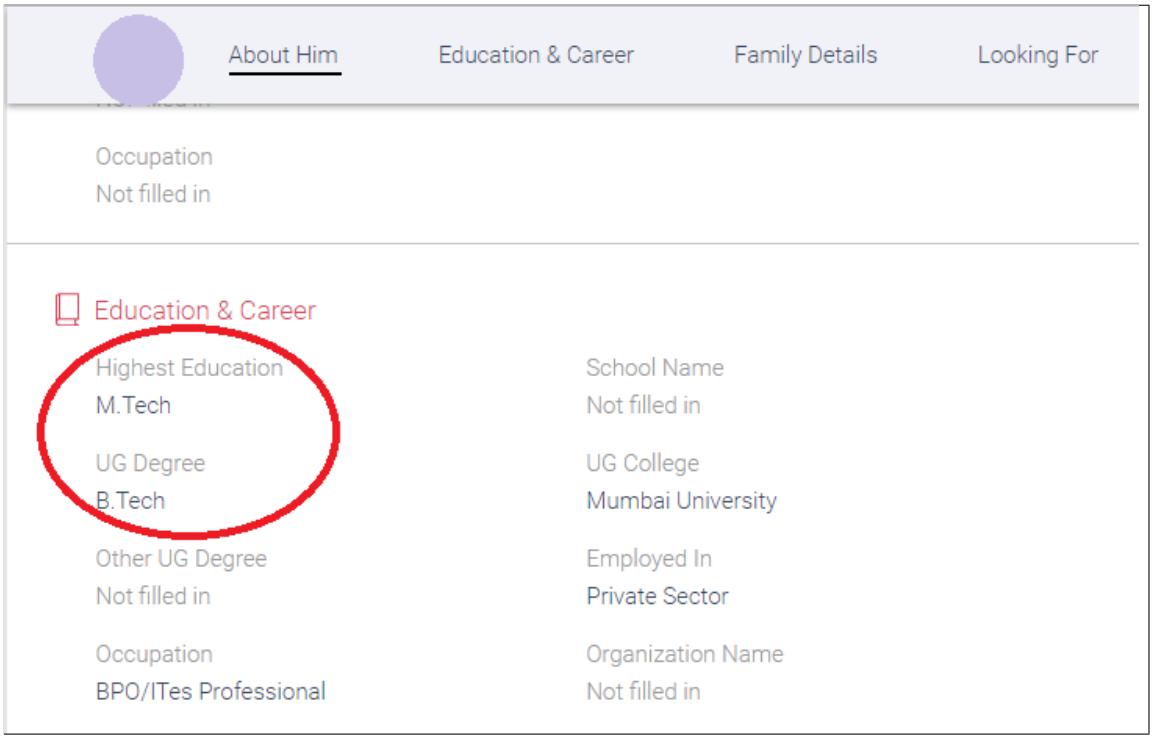

Figure 2.3: Fake profile initially registers on the portal with B.Tech and M.Tech as UG degree and highest education respectively.

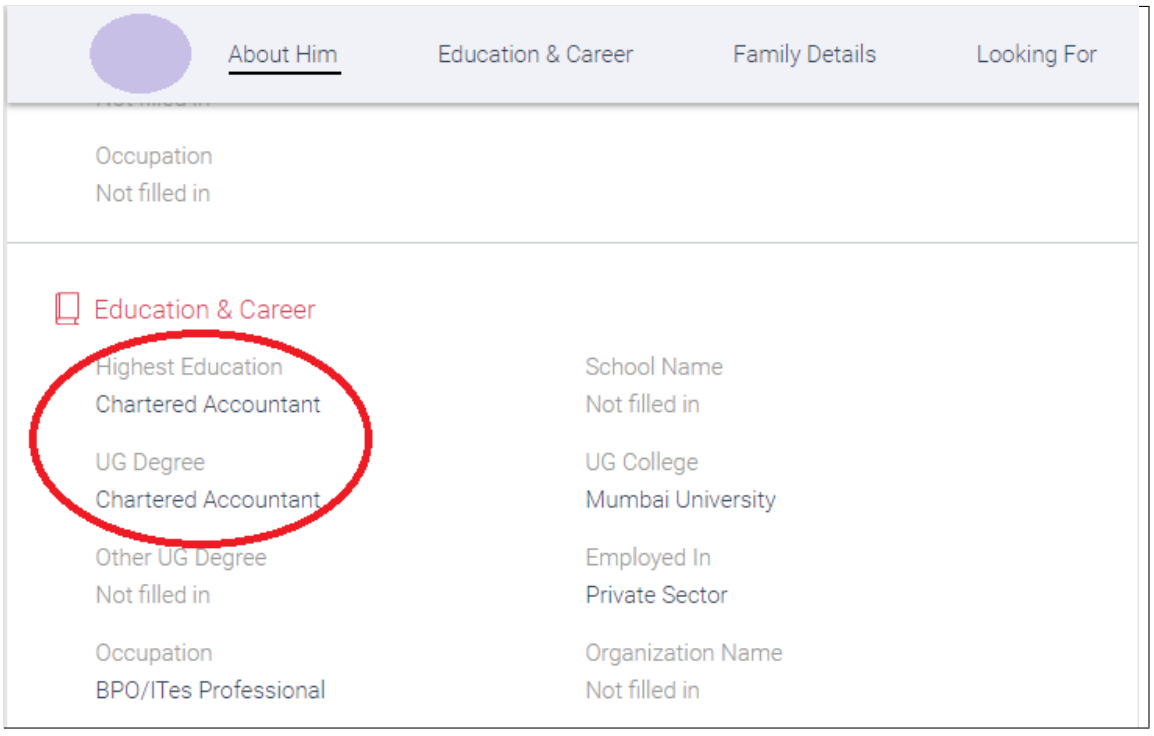

Figure 2.4: After few days, fake profile changes both UG degree and highest education fields to Charted accountant. This is quite suspicious activity.

2.2 Research Aim

In this thesis, we aim to address the following issues:

- On online matrimony, we try to find out the characteristics reflected by a fake profile
- We then compare these characteristics from those shown by genuine profiles on the platform
- We try to build a machine/deep learning based system which can be used in identifying fake profiles on the platform

To the best of our knowledge, this problem of detecting fake profiles on matrimony has not been delved into. This work is in collaboration with InfoEdge India Ltd. towards finding a machine learning based solution to this problem. To dig into this, we have chosen a use case of India‘s leading matrimony site

To detect fake profiles, we propose an anomaly detection method based on autoencoders, for which the input feature vector is decided judicially. We describe the feature vector in the coming chapters. The proposed framework will save many man hours that were used to spend in manually detecting fake profiles.

Chapter 3

Related Work

Online matrimony is a hub for the users searching for their desired life partner. Unfortunately, it also attracts users with malicious intentions to come on portal and spam peer users. Apart from Online matrimony, the trend of online spamming is prominent on other web forums like dating sites. [5] conduct study of scams/frauds done on an online dating website. They also proposed a taxonomy for different types of scammers present on an online dating website. [7] claims that bots are responsible for 51.8% traffic on online dating sites. They also conclude that these bots are very hard to differentiate from genuine users. As the functionality of online dating sites and online matrimony are quite similar, thus the existence of fake users is also a serious problem in matrimony domain [19]. Thus there is a need to build a machine learning based system which can detect such fake [1] users on online matrimony.

3.1 Detecting Fake Users

Different researches have contributed to detect fake users across different web domains. Hussain et al. [11] suggest a machine learning based framework to detect spam users in location based social networks (LBSN). The seriousness of this problem has also motivated researchers to identify spammer groups on social networking site like facebook. Hsu et al. [14] tried to detect spammers who have created Facebook groups to spread misinformation. They have included the relationship between members and characteristic of their activities in the feature set. They then trained a support vector machine to detect spammer groups on the portal. [25] also propose a machine learning system which detects fake user accounts on Facebook. Based on user profile activities and interaction with other users, they have developed a feature set which when fed into a machine learning classifier produces 79% accuracy. Adikari et al. [1] have detected fake profiles on the Linkedin dataset using machine learning algorithm. They claim that support vector machine with polynomial kernel outperforms other classifiers on the features captured from the Linkedin dataset.

[1] Fake and Spam are closely related terms. Thus in this work, we have used these terms interchangeably.

However, these proposed approaches are only specific to their domain of study. Due to huge difference in user profile attributes and user actions, the same work cannot be extended to detect Fake/Spam profiles on other domains like matrimony. Fake profile detection still remains an unsolved problem in matrimony domain.

3.2 Detecting Fake Content

One part of research in the area of spam detection has concentrated to detect spam content on web and social media. Kumari et al. [12] targets to classify user tweets as spam or not spam. They tried to detect tweets containing malicious links, fraudulent reviews to be spam and others as non-spam. Ghai et al. [13] tried to detect fake reviews on online review portal. They proposed a review processing method by which they used to detect reviews which have a higher degree of variation from other reviews and thus are probable spam content. Salminen et al. [9] address the problem of detecting hate/toxic content on the online platform. They first created a granular taxonomy for hateful comments on an online portal. Then they applied machine learning models like Decision tree, Logistic regression, Linear SVM, etc to detect such comments on the online platform. They claimed that Linear SVM has outperformed other models by achieving an average F1 score of 0.79. [10] propose a semi-supervised approach for opinion spam detection. They propose a framework where feature vectors of review, reviewer, and product are learned from existing techniques. They then applied classifiers on learned features to identify spamming. [4] targets to detect fake news articles on Twitter and tried to test different sets of classifiers. They were able to achieve 0.9 F1 score despite inaccurate annotated tweets. [16] have labelled 6.5 million spam tweets on twitter and extracted 12 lightweight features in order to facilitate research on spam detection on twitter. Moreover, they have experimented with some machine learning algorithms on the same dataset. [2] has even studied fake opinion problem on a well-known web forum of Taiwan. However, these works could not directly claim that if a user posting a spam content is a potential spammer or not.

3.3 Feature Engineering Approaches

Many papers propose novel feature designing method to detect fake profiles on web platforms. Green et al. [3] proposed edit based features to distinguish between a spammer and a genuine user on Wikipedia. Their edit feature set included edit size based features, time based features and link based features. The edit pattern of a user was learned by training XGBoost classifier. [13] incorporated both behaviour based and content based features in the machine learning model to detect social spammers on Twitter. Behaviour features include the number of user followers, number of followees, fraction of followers per followees, number of tweets posted by the user, etc. Other important features include minimum, maximum, median, average of time between user tweets and number of user tweets posted per day/week. Content based features include minimum, maximum, average and median of following attributes: proportion of the number of hashtags in

Types of Features	[3]	[13]	[22]	[6]	Proposed Features
User Behaviour Features		✓	✓		✓
User Edit Pattern	✓				✓
User Profile Features			✓	✓	✓
Affinity Features					✓

Table 3.1: Comparison of proposed features with the existing feature engineering techniques. Proposed feature set covers wide range of user's characteristics to predict him/her as fake.

a tweet, proportion of URLs per word, number of users mentioned in a tweet, etc. Behaviour features along with content based features were used in the classifier to predict if a profile is a spammer or a genuine user. [22] incorporated graph based features, user behaviour based features and content based features in the model to detect spammers on twitter. They then selected top k features using different methods like change of mean square, information gain and Relief-F techniques and noticed the improvement in the performance of the model. They report that reputation of account (Graph based feature), average length of tweet (average mention per tweet), average mention per tweet (Content based feature), age of account (Behaviour based feature) and the average time between posts (Behaviour based feature) to be most important features. [6] computed bag of words representation of user's messages and combined them with some expert features to detect spam users on social networking sites. These expert features include sex, age of both sender and receiver, user's account lifetime, month of birth, etc. They learned different classifiers like naive Bayes, logistic regression and SVM on the computed features.

Most of the research for spammer detection use content based features from the post written by a user on social media platforms. However, online matrimony does not provide an interface for its users to post content, thus features/methodology used in existing works cannot be directly applied to solve fake profile detection problem on matrimony. Moreover, on the matrimonial website, people have a specific mindset to interact/like users who belong to similar socio-economic attributes like age, income, caste, etc. Thus the user behaviour trend on matrimony is quite different from other web domains. Therefore behavioural spam detection work on other platforms cannot be extended in matrimony domain.

In this thesis, we propose a unique way of designing raw features capturing user's edit and behaviour pattern along with profile and affinity information. Affinity features capture how a user makes specific choices while sending interests to other users. We also realized that consolidated features capturing behaviour and edit heterogeneity can get diluted in long time windows. Thus we capture these features in smaller time windows (disjoint) to monitor user's activity/heterogeneity at much granular level. Table 3.1 compares the proposed feature set with existing feature designing techniques.

3.4 Machine Learning Techniques Used

Once feature extraction is completed, the next step which accounts for detecting spammers is the machine learning model that researchers have used. Spammer detection problem can be modeled as an anomaly detection task. Sabokrou et al. [8] compared the efficiency of Autoencoder, a non-linear dimensionality reduction technique with other existing reduction techniques like linear PCA and kernel PCA. This paper demonstrates that autoencoder can detect anomalies in data that even PCA fails to capture. Moreover, Denoising autoencoder can be used to further increase accuracy. Another advantage of using denoising autoencoder is that it does not involve computationally expensive operations as involved in kernel PCA. [15] performed anomaly detection in videos using Autoencoder approach. They proposed anomaly detection technique based on the reconstruction error of Autoencoder. They trained the Autoencoder network on normal video patches and then tested on unseen data. It was observed that an anomaly patch had a higher reconstruction error than that in case of the normal patch.

Autoencoder is also a widely used algorithm in spam detection problems. Mi G. et al [23] shows that the stacked Autoencoder outperform traditional algorithms like naive Bayes, decision tree, random forests, etc while detecting spam on different social media platforms. Castellini et al. [24] used Autoencoders in detecting fake profiles on Twitter. In this work, we explore how proposed feature set can be used with autoencoders to detect fake users on online matrimony.

Chapter 4

Contributions

In this thesis, we address the following issues:

- We try to do a qualitative analysis of both reported and unreported users on online matrimony.
- On matrimony website, we study how the heterogeneity in user behaviour can help us to distinguish between a fake profile and a genuine one.
- We also delve into the edit history of a user profile and try to observe the contrasting patterns between a fake profile and a genuine profile.
- We pinpoint the inconsistency in profile attributes of fake users and derive its importance in predicting a user as fake.
- Based on these important patterns, we propose a deep learning based approach to identify potential fake profiles on online matrimony.

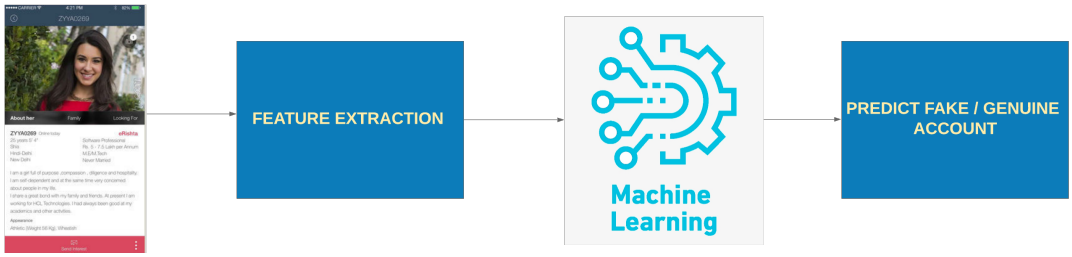

Figure 4.1: Contribution: Build a machine learning system which predicts if an account on matrimony platform is fake or genuine.

In this thesis, we first report the pattern differences between fake profiles and benign users. These differences include the pattern of sending interests messages to other users, the edit history of a particular user profile and the consistency between different profile attributes. These differences account for demarcating features which are fed into the deep learning model to identify fake profiles. To the best of our knowledge, this is the first study to detect fake profiles on matrimony.

Chapter 5

Analysis of Reported Users on Online Matrimony

On matrimony website, users report against peer users if they found them suspicious. They report based on behaviour and profile inconsistencies existing in other's profile. These reported profiles are then manually scrutinized to detect if they are fake or not [18]. This chapter tries to do an analysis of such reported/suspicious profiles.

5.1 Reported profiles

There exist two types of reported profiles: profiles reported by peer users and profiles reported by the system.

Following is the description of both types of reported profiles:

- User reported profiles: All users on the platform have the facility to report against the profile which they find having any malicious intentions. Malicious characteristics include using abusive language, demanding money, giving threats, etc. Every day many users get reported by others based on such factors and are manually scrutinized. After scrutiny, if a user profile is found as fake, then that profile is immediately deleted from the platform.
- System reported profiles: There is a system which reports profiles sending interest messages outside their specified religion, age and marital status preferences. However, this system only considers few attributes to report against profiles. These rule based systems are not confident enough to say if a profile is fake or not. Thus there is a need to build a machine learning based system for detecting fake profiles on the platform.

We analyze both types of profiles in the next sections of this chapter.

5.2 Detailed Descriptive Analysis on Reported Users

In this section, we aim to conduct a descriptive analysis of reported users present on the portal. To cater this objective, we found out the answers of following questions from the data of reported and unreported users. The first question focus to uncover the trend of the system reported and unreported profiles. Rest of the questions do the same for user reported and unreported profiles.

1. ***Question: What is the distribution of Total Score[1] given by system?***
   The frequency of users having score 0 is much higher than those having a score between 1-3. Figure 5.1 depicts that on matrimony website, we have an abundance of genuine profiles but very less suspicious ones. Although, these small number of suspicious users can be proved detrimental to the platform.

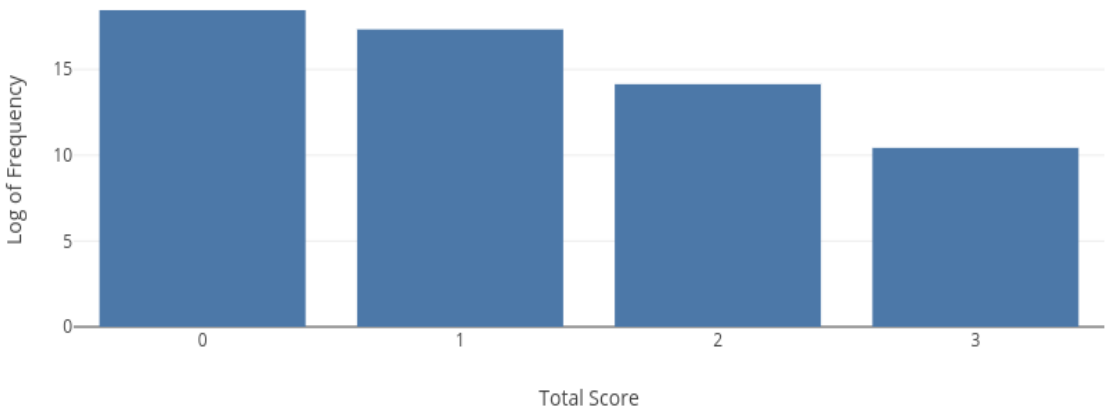

Figure 5.1: Majority of users on matrimony websites have zero score which means they have shown clean behaviour on the portal. This accounts for the fact that genuine profiles are in abundance as compared to suspicious profiles on matrimony portal.

2. ***Question: How many times each reported profile is reported by different users?***
   On online matrimony, each user can be reported multiple times by different other users. As per domain experts, even if a profile is reported around 3-5 times by different users, then it is a serious implication that the user is suspicious. Therefore, we analyzed how many times each reported user is reported by his/her peers. We found that most of the users are reported between 1-18 times. Moreover, there are a reasonable number of users which are reported between 18-35 and 35-52 number of times. This trend clearly shows us that reported users are actively doing spam on the platform and misbehaving with multiple users.

[1]System assess the level of suspicion of a user and rate him/her on a scale of 0 - 3. It is referred to as Total Score for a profile. 0 score being Not Suspicious and 3 being Very Suspicious.

Reasons of Reporting	Percentage of Cases
Looks like Fake profile	37.8
Inappropriate content	14.52
Already Married/Engaged	9.30
Incorrect details/photo	5.34
User has no intent to marry	4.68
One or more of profile details are incorrect	4.49
User is stalking me with messages/calls	3.96
User is not picking up phone calls	3.45
Photo does not belong to the person	2.92
Duplicate Profile	2.51
User is asking for money	1.96
Other Reasons	< 10

Table 5.1: Top reasons of reporting against a user along with their percentage of cases. Most frequently used reason is "Looks like Fake profile"

This also validates our claim that spam is a serious problem on matrimony which is affecting user experience on the site as well as matrimony business. Thus there is a need to build a feasible solution to detect such spam users on the platform.

3. ***Question: What are the top reasons used while reporting by peer users?***
   If a user reports against others, then he/she has to select the reason for reporting from the drop-down list. Table 5.1 illustrates the top reasons selected along with their percentage of cases. The most commonly used reasons are "Looks like a fake profile", "Inappropriate Content" and "Already Married/Engaged", "Incorrect details/photo".

4. ***Question: What is the month wise trend of reporting users?***
   We calculated discrete month wise frequency of reported users. We infer that the number of reported users increased from July 2016 to April 2017 but decreased from April 2017 to January 2018. A similar trend was observed while calculating month wise percentage of reported to active users. We noticed that the percentage of reported to active users in each month is really less[2].

5. ***Question: In a day, what is the distribution of interests sent by Unreported users?***
   We randomly picked 1000 unreported users and monitored their activity on the platform. According to our analysis, unreported users send maximum interests messages/initiates in two or three peak hours of the day and few in remaining hours. Figure 5.2 shows the plot between hour of the day and frequency of initiates.

[2]The exact data is confidential and can not be disclosed.

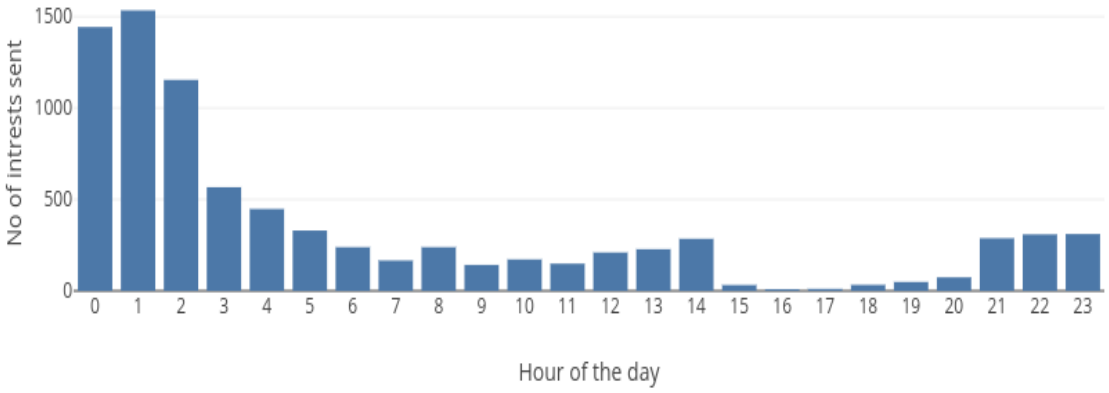

Figure 5.2: Most of the interests sent by unreported users on matrimony website are concentrated in the range of first 3 hours.

6. ***Question: In a day, what is the distribution of interests sent by reported users?*** We randomly picked 1000 reported users and monitored their activity on the platform. According to our analysis, reported users generally send maximum interests in specific time hours only. However, the frequency of interests sent by them in those specific hours is really high. One important observation includes that the total interests sent by 1000 reported users in a day clearly exceeds that of unreported users. Refer to Fig 5.3 for the same.

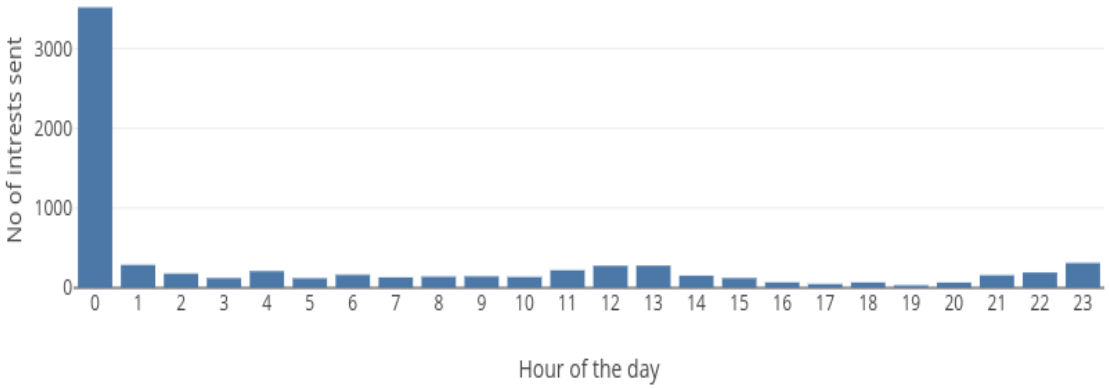

Figure 5.3: Interests sent by reported users on matrimony website are highly skewed towards a specific hour of the day.

In this chapter, we did a descriptive analysis on the data of reported and unreported users. This analysis helps us to get insights about the following aspects of reported and unreported users:

1. There is an abundance of genuine profiles as compared to system reported profiles.
2. A large bunch of user reported profiles have been reported multiple times by different users.
3. We also looked into top reasons of peer reporting.
4. We saw that the percentage of user reported profiles to active profiles is really less in each month.
5. We infer that unreported profiles send most of the interests messages in limited hour slot of a day. Whereas reported users remain active and send most interests messages in different time slots to spam more number of people.

Chapter 6

Uncovering Characteristics of Genuine and Fake Profiles

On matrimony website, if fake and genuine profiles coexist, then there has to be a unique pattern of fake profiles which make them stand out of the genuine ones. This chapter tries to find such patterns which can prove beneficial while predicting through a machine learning model. We try to dig into these patterns in the coming sections of this chapter.

6.1 Unboxing Patterns in User Interactions

The overall behaviour of a user includes how that user is interacting with peer users through certain actions. A user can interact with others through actions like sending interests, acceptances and rejects. Following is the significance of these actions in detail:

- **Interest Message**: Whenever a person likes somebody's profile, then he/she can send interest message to the other person. This message shows that the sender is impressed by the receiver's profile and is interested in marrying him/her.
- **Accept Message**: If the receiver of interest also likes the sender's profile, then he/she can send accept message to the sender.
- **Reject Message**: If the receiver of interest does not like the sender's profile, then he/she can send reject message to the sender.

We found the distinction in the pattern of sending interests messages by fake profiles. To illustrate the same, we have plotted the proportions of interest/initiates sent(to different categories of an attribute) by both genuine and fake profiles. Based on this, we compare the characteristics of both users. Following is attribute wise analysis of user interaction through interest message:

1. **Social Attribute: Caste**
   We took a random sample of 100 genuine and 100 fake profiles belonging to Aggarwal caste. We observed that genuine profiles generally send interest messages/initiates to caste categories which are similar to user's own caste. While sending interest messages, even if they deviate then the proportion of interests sent to other categories are very small. Whereas, fake profiles spam variety of users by sending interests across different categories. Fig 6.1 shows the distribution of initiates sent by a fake and genuine profile to different caste categories.

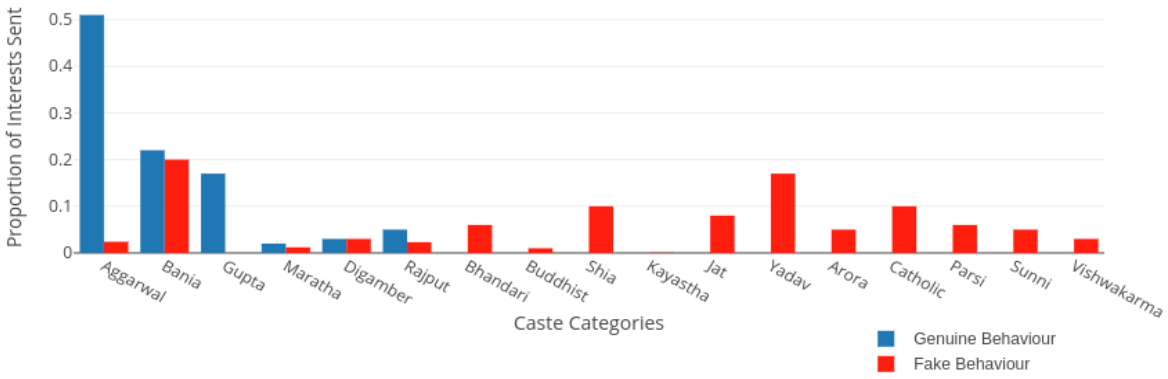

Figure 6.1: Genuine profiles generally send interests to specific categories of caste attribute. However fake profiles end up sending interests to multiple caste categories.

2. **Social Attribute: Marital Status**
   We took a random sample of 100 genuine and 100 fake profiles belonging to Non Married status. We observed that genuine profiles send most of the interest messages to same marital status categories. Fig 6.2 shows the distribution of interests sent by genuine profiles to marital status categories. Whereas, fake profiles spam variety of users by sending interests across different categories of marital status. Sometimes it is quite abnormal activity that the same user is interested in marrying non-married, divorcees and separated people. Fig 6.2 shows the distribution of initiates[1] sent by fake and genuine profiles to different marital status categories.

[1]Initiates and Interests technically mean the same. Thus we have used these words interchangeably.

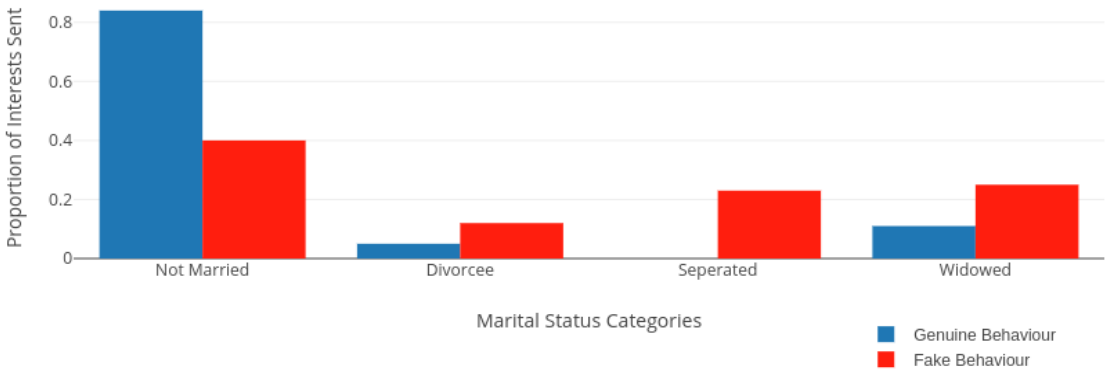

Figure 6.2: Genuine profiles send most of the interests to specific categories of marital status attribute. However fake profiles end up sending interests to multiple marital status categories.

3. **Social Attribute: Mother Tongue**

   We took a random sample of 100 genuine and 100 fake profiles belonging to Hindi-Delhi category. We observed that genuine profiles send most of the proportions of interest messages to selected mother tongue categories and very less proportions to other categories. Whereas, fake profiles send small proportions of interests to multiple categories. Fig 6.3 shows the distribution of interests sent by genuine and fake profiles to different mother tongue categories.

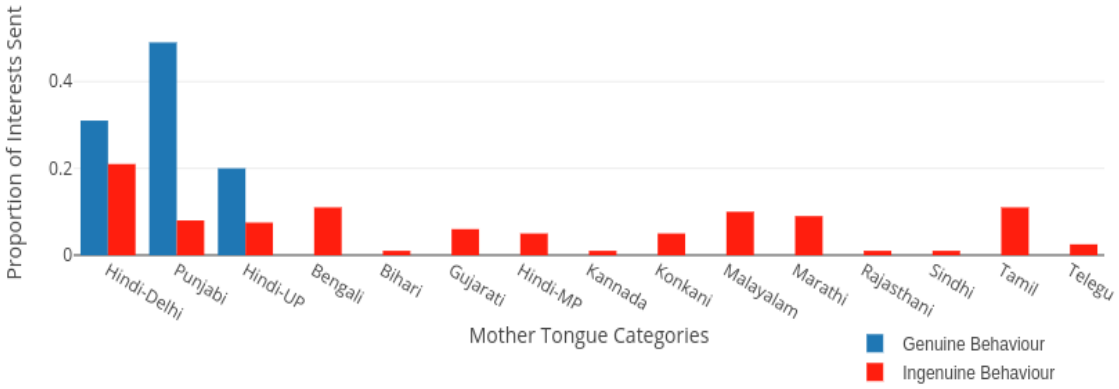

Figure 6.3: Genuine profiles generally send interests to specific categories of mother tongue attribute. However fake profiles end up sending interests to multiple mother tongue categories.

4. **Economic Attribute: Income**

   We took a random sample of 100 genuine and 100 fake profiles belonging to Rs 3-4 Lakh category. For the income attribute, we have observed that fake profiles exhibit more heterogeneity in sending interest messages as compared to genuine profiles. From figure 6.5, we can observe that some fake profiles send interests to low INR income but at the same time they target people having a high income in US dollars. This behaviour is quite abnormal.

Figure 6.4: Both fake and genuine profile generally shows more heterogeneity in sending interests as compared to that of genuine profiles. It is seen that many fake profile even target some users having high salary in US dollars.

In this section, we infer that a genuine profile generally sends interest messages to the category of users whom he/she is really interested in marrying. However, a fake profile tends to send interest messages to a variety of categories of users. There is a high probability that he/she is not interested in all of them. We call this pattern as **heterogeneity in sending interests/ behaviour abnormality**. According to domain experts, categories of many users(who are receivers of interest messages sent by fake profile)do not even lie in the desired partner attributes of that fake profile. Thus sending interest messages to a variety of users is one of the way adopted by fake profiles to spam multiple people and then contact them for money or other benefits. A similar pattern of behaviour heterogeneity is also validated on different platforms like Twitter [20, 21]

6.2 Unboxing Patterns in Editing Profile Information

We have found the distinction in the pattern of editing own profile attributes by fake profiles. To illustrate the same, we have plotted the time proportions spent(on different categories of an attribute) by a user in his/her lifetime. Based on this, we compare the characteristics of fake and genuine users. Following is attribute wise analysis of the user's edit pattern:

1. **Social Attribute: Caste**

   We took a random sample of 100 genuine and 100 fake profiles belonging to Aggarwal category. We observed that genuine profiles generally remain on one category of caste throughout their lifetime. However, fake profile keeps on editing their caste frequently. Figure 6.6 shows the time proportions spent by genuine and fake profiles on different categories of caste.

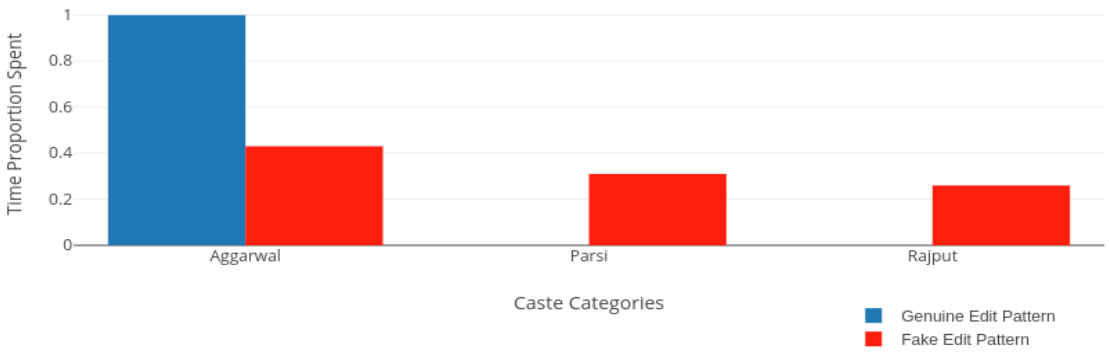

Figure 6.5: Genuine profiles have spent their full lifetime on one caste category. However fake profiles have spent different time proportions on different caste categories.

2. **Social Attribute: Mother Tongue**

   We took a random sample of 100 genuine and 100 fake profiles belonging to Hindi-UP category. We observed that genuine profiles do not frequently edit their mother tongue category. Even if they edit it, then the new category is quite similar to the old one. However, fake profiles keep on editing their mother tongue frequently. Figure 6.7 shows the time proportions spent by genuine and fake profiles (respectively) on different categories of mother tongue.

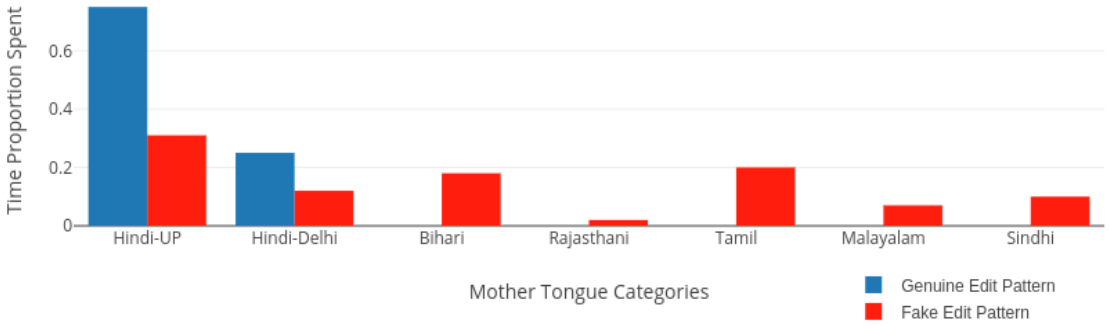

Figure 6.6: Genuine profiles have spent large proportions of their lifetime on one mother tongue category. However fake profiles have spent small time proportions on different mother tongue categories.

3. **Economic Attribute: Income**

   We took a random sample of 100 genuine and 100 fake profiles belonging to Rs 5-7.5 Lakh category. For the income attribute, it is observed that both genuine and fake profiles generally edit this attribute. However, it is noticed that for fake profiles, there is a high variation between the categories before and after an edit. Fig 6.8 shows time proportions spent on different income categories by both fake and genuine profiles. Here, fake profile shows a big jump in category Rs 25-35 Lakhs and Rs 50-70 Lakhs from category Rs 5-7.5 Lakhs and so forth. This accounts for abnormality in edit pattern on the behalf of the user. Such a pattern is not visible in a genuine profile.

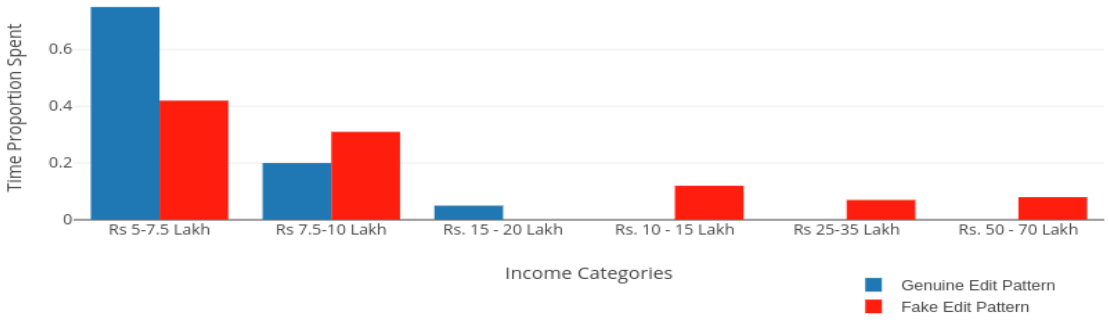

Figure 6.7: Both genuine and fake profiles edit income frequently. But the sudden transition between income categories by fake profile is sometimes suspicious.

In this section, we infer that a genuine profile generally sticks to one category throughout his/her lifetime. Even if he/she switches to other categories, the maximum time is spent on only one category. However, a fake profile tends to switch between multiple categories frequently. As

a result, fake profile spends a bit uniform time proportions on multiple categories. We call this pattern of fake profiles as **Inconsistent edits**. According to domain experts, fake profiles frequently make such edits so as to get different recommendations each time and then spam different variety of users on the platform. This edit pattern is similar to the update pattern of fake accounts found on Twitter [20] and Wikipedia platform [3].

6.3 Unboxing Patterns in Profile Attributes

We have encountered some fake profiles which also exhibit inconsistency within profile attributes specified by them. For instance: a 21 years old person is claiming to hold a Ph.D. degree. These details seem to be suspicious on the user's behalf. We refer to this characteristic of fake profiles as **Profile Inconsistency**. Figure 2.1 and 2.2 show the cases of profile inconsistency present in some fake accounts.

6.4 Inference

In this chapter we infer the following :

- Fake profile exhibit more heterogeneity in sending interest messages as compared to genuine profiles.
- Fake profiles also do inconsistent edits in their profile information while this may not be the pattern of genuine profiles.
- Fake profiles show the inconsistency in profile attributes which generally is not the trend in genuine profiles.
- Capturing interest heterogeneity, edit inconsistency and profile inconsistency in the form of features can help the machine learning model to predict better.

Chapter 7

Methodology Adopted

7.1 Feature Engineering

On online matrimony, a user profile posses different socio-economic, physical and professional attributes. These all are categorical attributes which are used to build proposed set of raw features to detect fake profiles. The proposed feature set captures user's behaviour, edit summary, profile information along with his affinity towards other categories. Moreover, behaviour and edit summary features were computed in different time buckets to provide a finer level of granularity.

The set of attributes used by us during feature engineering are: age, body type, caste, city, country, education, height, income, manglik, marital status, mother tongue, occupation, religion.

7.1.1 Dynamic length window

We captured interests heterogeneity and edit pattern of a user in different time windows. If the number of time windows taken is w, then the total interests sent by the user are divided in these w time windows such that each window contains an equal number of interests. Once these windows are fixed for a user, then heterogeneity and edit features are computed in each of the windows. Along with them, we also store the time duration and number of initiates sent as features in each window. As the length, start and end time of a window can vary from user to user, thus we refer these windows as dynamic windows.

7.1.2 Behavior Features

Section 6.1 showed the distribution of interests of a fake and a genuine profile for different categorical attributes. For an attribute like caste, religion, mother tongue, a Fake profile tries to send interest to most of the categories. On the other hand, a genuine profile tends to send interest only to those categories in which he is really interested in. This unique demarcation in pattern of sending interest becomes an important feature to judge if a user is authentic or not. So to capture user's behavior of sending interest in last 60 days of his activity we have designed behaviour features in the following manner.
If there are x windows namely W_1,W_2,W_3...W_x. An attribute, A has n categories namely C_1,C_2,C_3...C_n.
Let $I^k_{i\to cj}$ denote the number of interests sent by user U_i to category C_j in window W_k. Then behaviour feature of user U_i for category C_j in the window W_k is defined in the following way:

$$B^k_{i\to cj} = \frac{I^k_{i\to cj}}{\sum_{j=1}^{n} I^k_{i\to cj}}$$

In each time window, we store the frequency of interests sent by a user to different categories of an attribute . We then normalized this value by the total number of interest sent by him in the same time window.

Technically, we are trying to store the probability distribution of a user while sending interests to different categories. As storing the full information of interests could be computationally expensive, thus we stored the summary of interests sent by a user in 1 dimensional feature vector. Moreover, capturing the behaviour heterogeneity in single time window could dilute the quantities of proportions. Thus we came up with the idea of capturing the same information in non overlapping different time windows where the number of windows are same for all users.

7.1.3 Edit Features

Figures in section 6.2 shows the time proportions spent by a fake profile and a genuine profile on different categorical attributes. It is evident that fake profile edit some crucial attributes like caste and mother tongue which ideally should not be changed in a user's lifetime. Moreover, the time difference between consecutive edits of a fake profile(for different categorical attributes) is very less whereas genuine profiles generally stick to one category for longer time. To capture this distinction in the edit pattern, we computed edit features in the following way.
If there are x windows namely W_1,W_2,W_3...W_x. An attribute 'A' has n categories namely. C_1,C_2,C_3...C_n. Let S^k and E^k denote the start and end time stamp of window W_k respectively . Moreover, $S^k_{i\to cj}$ denote the time stamp when User U_i changed his category to C_i in window W_k. Similarly $E^k_{i\to cj}$ denote the time stamp when User U_i left category C_i in window W_k

Edit feature $Ed^k_{i\to cj}$ of user U_i with respect to category C_j in the window W_k is defined as:

$$Ed^k_{i\to cj} = \frac{(E^k_{i\to cj} - S^k_{i\to cj})}{(E^k - S^k)}$$

In each time window, we are storing edit snapshot of user's profile attributes. Edit snapshots are characterized by the amount of time a user has spent on a particular category. We then normalized these time quantities by the total time duration of that window.

Storing the full information of edits could be computationally expensive, thus we stored the summary of edits done by a user in 1 dimensional feature vector. Capturing the inconsistency edit pattern in single time window could dilute the quantities of proportions. Thus we came up with the idea of capturing the same information in non overlapping different time windows where the number of windows are same for all users

7.1.4 Affinity Features

An Affinity score between two categories (C_i,C_j) is the likelihood score of a person having category C_i to send interests to user having category C_j. If an attribute 'A' has n categories namely C_1,C_2,C_3...C_n. Let $U_1^{C_i}$,$U_2^{C_i}$,$U_3^{C_i}$...$U_m^{C_i}$ be the users present in category C_i. Also, $I^m_{ci\to cj}$ denotes the number of interests sent by $U_m^{C_i}$ (having C_i category) to C_j category. Then $A_{ci\to cj}$, affinity feature between category C_i and category C_j is defined in the following way:

$$A_{ci\to cj} = \frac{\sum_{k=1}^{m} I^k_{ci\to cj}}{\sum_{j=1}^{n}\sum_{k=1}^{m} I^k_{ci\to cj}}$$

A vector storing affinity scores of a user towards different categories represent his expected probability distribution of interests's frequencies towards other categories. These affinity scores between categories (C_i,C_j) were statistically computed based on proportion of interests sent by category C_i to category C_j in the past. Thus affinity scores provide the real picture of how a user having category C_i is expected to behave on the platform. On matrimony platform, a user tend to send more interest to categories which are similar to his own category. Thus similar categories have higher affinity scores as compared to completely dissimilar categories.

In the feature vector of a user having category C_i, we append one dimensional vector storing affinity scores of C_i with all other categories C_j. If total number of categories of an attribute are n, then the length of affinity vector for that attribute is also n . This one dimensional affinity vector is computed along each attribute.

The behaviour feature just captures the heterogeneity of a user in sending interest. However, if a user shows heterogeneity within similar categories , then he can be a genuine profile. Thus incorporating affinity vector with behaviour vector gives the model, an information about the similarities between the categories of user and the person to whom user is initiating. Resulting

model predicts fake only if the user is showing heterogeneous behaviour in dissimilar categories. This in turn helps the model to reduce the number of false positives in predicting spammers and thus improves the precision score.

7.1.5 Profile Features

A unique characteristic of a fake profile is its inconsistency within profile attributes. There have been cases when a user claims to hold a Ph.D. degree at the age of 21 years. This accounts for profile inconsistency along the education and age attributes. Also, there exist inconsistency between the user's own category and the category to whom he/she is sending interests. For instance, if a non-married user is sending interests to the divorced users, then it can be an abnormal activity on the behalf of the user, thus he/she can be a potential spammer. To capture these inconsistencies across different categorical attributes, we incorporated one hot vector of the user's profile attributes. By profile and behaviour features together, the model learns that a user having particular category c_1 of an attribute X is sending interest to users of category c_2 of the same attribute X.

7.2 Feature Selection

From the pool of features, there was the need to select important features that can enable the model to easily distinguish between a Fake and a genuine profile. The measure of entropy served this purpose for us. For each attribute we calculated the entropy on data of fake profiles, let's denote it as SE and the entropy on data of genuine profiles, let's denote it as CE. We then define the Diff as the difference metric in the following manner:

$$Diff = \frac{(SE - CE)}{SE}$$

The higher the value of Diff we get for an attribute, the more that attribute is able to discriminate between a Fake and a clean profile. Thus we chose only those attributes for which we have got the value of Diff metric to be higher than a threshold.

7.3 Training Using Autoencoder

We had an abundance of data of genuine profiles as compared to that of Fake ones. Thus the problem of class imbalance can arise while taking it as two class supervised learning problem. Our idea is to learn the distribution of genuine profiles as their behavior and edit information exhibit a pattern. Thus we trained Autoencoder on the feature vector of genuine profiles as one class learning process where reconstruction error (mean square error) was used to distinguish between a Fake and a genuine profile.

We had a pool of around 5,40,737 users who were active in the last 2 months. In this pool of data, we had some users who were marked as Fake either by peer users or by rule based system. After removing all such suspicious users from the total pool of data, we get the user profiles which have shown genuine behaviour, profile and edit pattern on the portal. We computed feature vector of these profiles and trained stacked Autoencoder containing 3 encoding layers and 3 decoding layers. In the Autoencoder, we used used relu activation function in the hidden layers and sigmoid in the output layer. We also used adam optimizer and glorat normal for initializing weights.

During training, Autoencoder learns the pattern that genuine profiles exhibit, thus any Fake profile deviating from the learnt pattern will show a high reconstruction error. Whereas, a genuine profile showing a similar pattern will produce low reconstruction error.

We then selected a threshold on reconstruction error using the elbow of validation loss curve. Users above this threshold value are predicted as Fake profiles and below ones are predicted as genuine profiles.

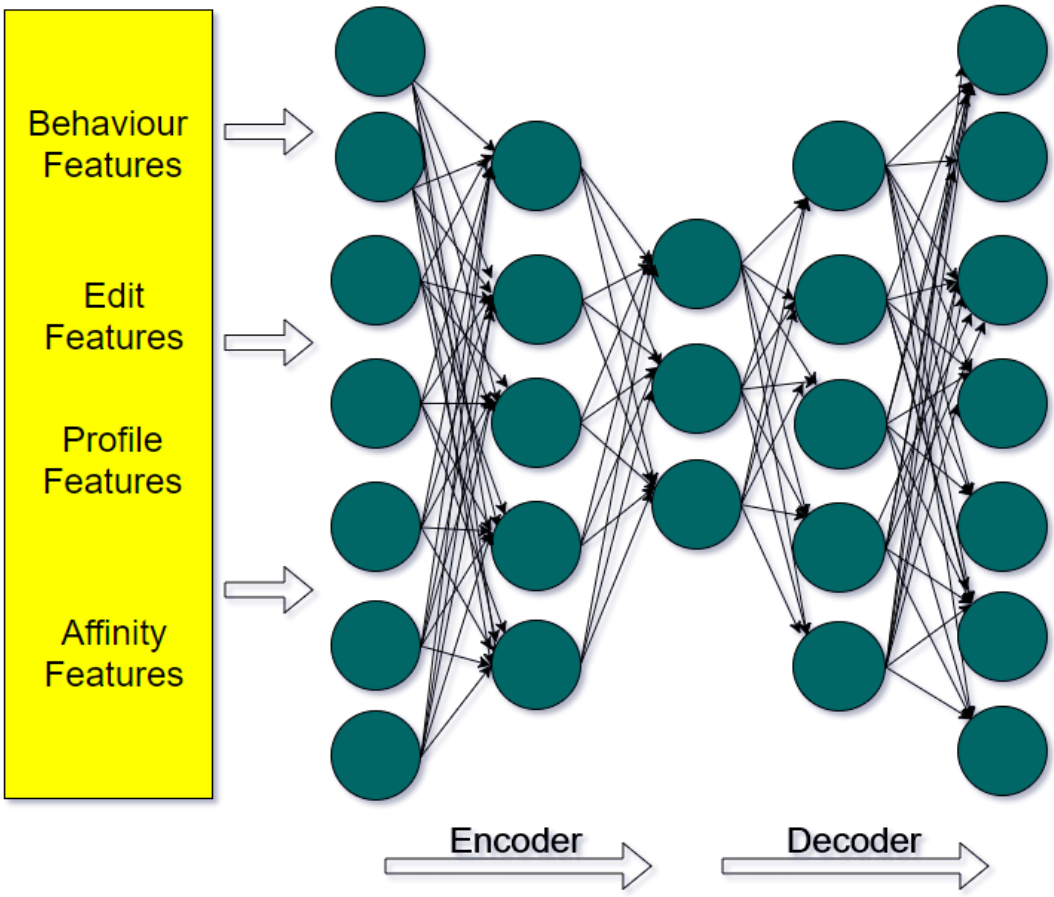

Figure 7.1: Autoencoder model included 3 encoding and 3 decoding layers. We have used relu activation function on the hidden units and sigmoid activation on output units.

Chapter 8

Experimental Results

8.1 Using Behaviour Features

As per domain experts, the initial 8 days user behaviour on the portal is good enough time to predict him as a genuine/fake profile. Thus we trained autoencoder model on features capturing the first 8 days behaviour on the platform. We then tested the model on the offline dataset and report the following results :
We observed good results on offline test data. Table 8.1 shows the confusion matrix of the same. However, when this model was deployed on the platform, results on online test data were not encouraging. Table 8.2 and 8.3 shows true positives and false negatives cases respectively on online test data (along with their reasons for spam).

In Table 8.2, the highlighted text shows the reasons which are related to user's abnormal behaviour trend. This shows that the model is able to capture such abnormal behaviour cases. Moreover, Table 8.3 shows the cases when the model predicted some profiles as genuine but they were actually Spam/Fake. After looking at the reasons of spam, we can see that most of such cases are the ones who have done multiple edits in their profile or have kept inconsistent profile information. Thus we infer that only behaviour feature is not capable to correctly classify a profile as fake/genuine. We realize the need of incorporating edit and profile features while training so that model does not miss on spam/fake cases mentioned in table 8.3.

One major drawback of this approach is that the user has to be atleast 8 days old on the portal before being tested from the model. Thus a fake profile can exploit the platform(in less than 8 days) before being detected by the algorithm. To overcome this, we came up with the concept of

Confusion Matrix	Predicted Fake	Predicted Genuine
Actual Fake	2953	852
Actual Genuine	168	17799

Table 8.1: Confusion matrix obtained after testing behaviour model on the offline dataset.

Reasons for Fake
One more profile with different marital status has been observed for the member
In the about me section mentioned as marketing professional and in occupation mentioned police.
Initiating in all marital status, age group and income
Never married profile initiating with divorced profile and profiles from age 26-35
Seems spam based on the dubious trend and information mentioned in the profile

Table 8.2: Reasons associated with fake profile which are present in true positive cases

Reasons for Fake
Suspicious attributes. Profile with parents expired, IT professional with 200,001 US dollars
Member has changed his salary thrice in just 6 days of registration
Changed his contact no. 4-5 times in single day
22 years old but sending interest upto 26 years, divorced profiles.

Table 8.3: Reasons associated with fake profiles which are present in False Negative Cases

dynamic length windows (section 7.1.1). These dynamic windows enable us to test users on each day when they are active. This helps in early and pro active identification of fake users on the platform. Exlj98periments mentioned in next sections are based on such dynamic windows.

8.2 Incorporating Behaviour, Edit and Profile Features

As in the previous section, we observed that only behaviour features are not sufficient to classify a person as Fake/genuine. Thus in order to capture the overall characteristic of a profile, we incorporated behaviour, edit and profile based features in our model. We then experimented these features for a different number of dynamic time windows as discussed in section 7.1.1. If the number of windows taken for an experiment are W, then we will divide the last 60 days activity of each user in these W time buckets. This concept of taking W time windows will help in detecting abnormality in behaviour, profile and edit pattern of a user at finer granularity levels.

Table 8.4 shows the precision, recall and accuracy scores obtained for different window sizes. It shows that both precision and recall scores significantly improve when we take a single window. We refer it as the best model till now.

We then applied feature selection method on the best model obtained (Using a single window). Table 8.5 shows the increase in precision, recall and accuracy scores of the model after applying feature selection method. It shows that the model performs better after selecting the features having a higher difference in entropy values across Spam/Fake and Genuine user's data.

Method	Precision	Recall	Accuracy
Using Five windows	0.170	0.510	0.8830
Using Two windows	0.230	0.780	0.8977
Using one window	**0.266**	**0.866**	**0.8972**

Table 8.4: Comparing performance of models while taking different time windows

Method	Precision	Recall	Accuracy
Using one window	0.266	0.866	0.8972
Using one window + Feature Selection	**0.269**	**0.894**	**0.9083**

Table 8.5: Best model before and after feature selection

8.3 Incorporating Behaviour, Edit, Profile and Affinity features

Even after conducting previous experiments, the number of false positives did not decrease significantly. We observed that such false positive cases include profiles who have shown some heterogeneity in behaviour but this pattern of heterogeneity is also common in their community. Thus to reduce such cases, we have incorporated affinity features in the existing model. These affinity features (discussed in section 7.1.4) describe how a user of a particular community generally behaves with users across other communities. Affinity scores when incorporated with behaviour features compare between how a user is expected to behave and how he/she actually behaves on the platform.

This exercise has led to a significant increase in precision score (because false positive decreases). Table 8.6 contains improved results.

Precision	Recall	Accuracy
0.341	0.902	0.9176

Table 8.6: Results after using full feature Vector

Method	Precision	Recall	Accuracy
Entropy Features + Autoencoder	0.090	0.240	0.8700
Entropy and Affinity Features + KDE	0.087	0.495	0.7380
Entropy Linear Sum + Autoencoder	0.070	0.661	0.8971
Proposed Features + Autoencoder	**0.341**	**0.902**	**0.9176**

Table 8.7: Comparing all models with baselines

8.4 Comparison with Baseline

Ideally, processed features like attribute wise entropy along with affinity scores should work satisfactorily. However, when these features are combined and trained on different sets of algorithms (like kernel density estimation and autoencoders) they could not produce desired results. On the other hand, when the proposed set of features (in section 8.3) are trained using autoencoders, then it outperforms other methods as shown in Table 8.7.

Chapter 9

Real World Impact

Before this work, there was no automated process which could detect fake users on matrimony portal. Although, there exists a rule-based system but it could not detect fake users with high confidence. Thus the proposed machine learning framework is highly desirable for removing fake users from the online matrimony portal. It saves many man hours which were used to manually scrutinize profiles every day and then delete them.

The proposed machine learning system is being deployed on the web portal. Everyday algorithm test on the active profiles present on the website. It then ranks these profiles in decreasing order of reconstruction error; profile having the highest error being the most probable spammer. The list containing top k scores are then returned to the operations team. Operations team scrutinize these profiles and delete fake profiles.

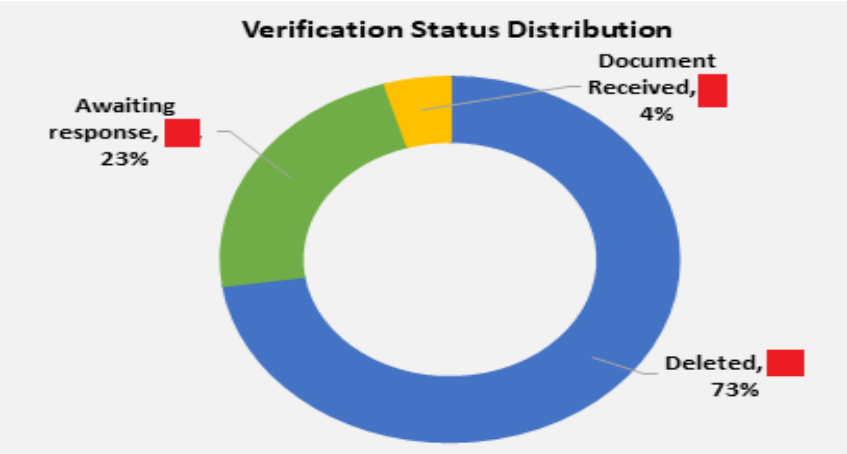

Figure 9.1: 73% of profiles in verification status are found as fake and deleted from the matrimony platform. Only 23% users need to be more verified and are their response is awaited. Only 4% profiles are detected as genuine profiles. This trend shows that profiles returned by model are mostly fake.

Before this work, lakhs of user profiles were manually scrutenized to find spammers. But after this model was deployed, only the list of most probable fake profiles was returned by the model for manual scrutiny . This practice substantially reduced their set for manual scrutiny. After manual scrutiny, some profiles are identified as sure shot spammers, whereas some profiles fall

under verification required status. Profiles falling in verification required category are asked to present their identification documents. If they don't turn up, then these profiles are deleted from the portal. Figure 9.1 shows the verification status distribution of profiles in April, 2019. It clearly shows that 73% profiles did not turn up and are probable spammers. We can also see that in one month, only 4% profiles's documents were correctly verified.

Chapter 10

Conclusion, Limitations, Future Work

10.1 Conclusion

Fake profile detection is a challenging problem on online matrimony. A fake profile uses multiple ways to spam people on the portal. Thus it is difficult to detect them on the basis of only one characteristic. Multiple ways of spamming also lead to more exploitation done by such profiles to peer users.

In this thesis, we contributed towards finding out the unique characteristics of fake users. In chapter 5, we did a descriptive analysis of both reported and unreported users. We then studied the distinction in behaviour, edit pattern and profile information of genuine and fake users in chapter 6. Using this study, we incorporated behaviour and edit features using dynamic time windows. We extended this feature set by including profile and affinity scores. We then trained the autoencoder model which outperforms traditional feature engineering methods and solve Fake/Spam problem on online matrimony.

To the best of our knowledge, this is the first study done to detect fake profiles on online matrimony. This study can be used

10.2 Limitation

For training autoencoder, we could only get 5,40,737 genuine profiles. As the number of genuine profiles was less, thus model could learn a limited distribution of genuine profile behaviour. Thus any profile which is slightly deviating from the learned distribution was being predicted as a fake profile. This leads to an increase in false positive cases and less precision score.

We also could not explore the spamming done through chats/text being written during a conversation with peer users. This was because we could not get access to private conversations done between users on the platform.

10.3 Future Work

More efforts can be applied to increase the precision score of the existing model. One of the ways to achieve it is by incorporating more genuine samples in the training set. This will help the model to learn a more generalized genuine pattern and reduce false positive cases.

The proposed generalized behaviour, edit and profile can be used to detect fake profiles on other social media/web platforms. Moreover, if the text conversation's data is available. then this work can be extended to incorporate textual features in the existing pipeline of features.

This work focused more on feature engineering task than modeling different classifiers. Thus we can also try other machine/deep learning models which can be proved to be more accurate than the existing state of the art algorithms.

Bibliography

[1] Shalinda Adikari and Kaushik Dutta. Identifying fake profiles in linkedin. In 18th Pacific Asia Conference on Information Systems, PACIS 2014, Chengdu, China, June 24-28, 2014, page 278, 2014.

[2] Li-Chen Cheng, Judy C. R. Tseng, and Tsai-Yu Chung. Case study of fake web reviews. In ASONAM, 2017.

[3] Thomas Green and Francesca Spezzano. Spam users identification in wikipedia via editing behavior. In Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017, Montreal, Quebec, Canada, May 15-18, 2017, pages 532, 533, 534, 535.

[4] Stefan Helmstetter and Heiko Paulheim. Weakly supervised learning for fake news detection on twitter. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 274 to 277, 2018.

[5] JingMin Huang, Gianluca Stringhini, and Peng Yong. Quit playing games with my heart: Understanding online dating scams. In International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, pages 216âĂŞ236. Springer, 2015.

[6] Antonio Lupher, Cliff Engle, and Reynold Shi Xin. Feature selection and classification of spam on social networking sites. 2012.

[7] Tahora H. Nazer, Fred Morstatter, Gareth Tyson, and Huan Liu. A close look at tinder bots. 2017.

[8] Mayu Sakurada and Takehisa Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, Gold Coast, Australia, QLD, Australia, December 2, 2014, page 4, 2014.

[9] Joni Salminen, Hind Almerekhi, Milica Milenkovic, Soon-Gyo Jung, Jisun An, Haewoon Kwak, and Bernard J. Jansen. Anatomy of online hate: Developing a taxonomy and machine learning models for identifying and classifying hate in online news media. In Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018., pages 330âĂŞ339, 2018.

[10] Cennet Merve Yilmaz and Ahmet Onur Durahim. Spr2ep: A semi-supervised spam review detection framework. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 306âĂŞ313, 2018.

[11] Hussain A., Keshavamurthy B.N. (2019) AnalyzingâĂŸOnline Location-Based Social Networks for Malicious User Detection. In: Sa P., Bakshi S., Hatzilygeroudis I., Sahoo M. (eds) Recent Findings in Intelligent Computing Techniques. Advances in Intelligent Systems and Computing, vol 707. Springer, Singapore

[12] Vidya Kumari K.R., Kavitha C.R. (2019) Spam Detection Using Machine Learning in R. In Smys S., Bestak R., Chen JZ., Kotuliak I. (eds) International Conference on Computer Networks and Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies, vol 15. Springer, Singapore

[13] Ghai R., Kumar S., Pandey A.C. (2019) Spam Detection Using Rating and Review Processing Method. In: Panigrahi B., Trivedi M., Mishra K., Tiwari S., Singh P. (eds) Smart Innovations in Communication and Computational Sciences. Advances in Intelligent Systems and Computing, vol 670. Springer, Singapore

[14] Fu-Hau Hsu, Meng-Jia Yan, Kai-Wei Chang, Chih-Wen Ou, Hung-Min Sun. Itus: Behavior-based Spamming Group Detection on Facebook. In: Airiti Library 10.3966/199115992018082904006

[15] M. Sabokrou 1 ; M. Fathy 2 ; M. Hoseini. Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder. IET Digital library 10.1049/el.2016.0440

[16] Intuit forecast: 7.6 million people in on-demand economy by 2020. Accessed: April 12, 2018. http://math.tntech.edu/rafal/cliff11/index.html

[17] Pew Research Centre Survey: https://www.pewsocialtrends.org/2014/09/24/chapter-1-public-views-on-marriage/

[18] Fraudulent Cases: https://www.jeevansathi.com/static/page/fraudalert

[19] https://timesofindia.indiatimes.com/city/pune/matrimonial-fraud-on-the-rise-more-than-50-cases-registered-this-year/articleshow/60049950.cms

[20] Gurajala, S., White, J. S., Hudson, B., Voter, B. R., Matthews, J. N. (2016). Profile characteristics of fake Twitter accounts. Big Data Society. https://doi.org/10.1177/2053951716674236

[21] H. Shen and X. Liu, "Detecting Spammers on Twitter Based on Content and Social Interaction," 2015 International Conference on Network and Information Systems for Computers, Wuhan, 2015, pp. 413-417. doi: 10.1109/ICNISC.2015.82

[22] Herzallah, W., Faris, H., Adwan, O. (2018). Feature engineering for detecting spammers on Twitter: Modelling and analysis. Journal of Information Science, 44(2), 230âĂŞ247. https://doi.org/10.1177/0165551516684296

[23] Mi G., Gao Y., Tan Y. (2015) Apply Stacked Auto-Encoder to Spam Detection. In: Tan Y., Shi Y., Buarque F., Gelbukh A., Das S., Engelbrecht A. (eds) Advances in Swarm and Computational Intelligence. ICSI 2015. Lecture Notes in Computer Science, vol 9141. Springer, Cham

[24] Castellini, Jacopo Poggioni, Valentina Sorbi, Giulia. (2017). Fake Twitter followers detection by denoising autoencoder. 195-202. 10.1145/3106426.3106489.

[25] Gupta, Aditi Kaushal, Rishabh. (2017). Towards detecting fake user accounts in facebook. 1-6. 10.1109/ISEASP.2017.7976996.

[26] Lok Foundation Survey: https://www.lok-foundation.org/lok-survey-project/

[27] Desai, Sonalde, and Vanneman, Reeve. India Human Development Survey-II (IHDS-II), 2011-12. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2018-08-08. https://doi.org/10.3886/ICPSR36151.v6