Heriot-Watt University

April 2025 Student Submission

Detecting Fake Accounts on LinkedIn

Author:
Guilhem Vinet

Supervisor:
Marwan Fuad

A thesis submitted in fulfilment of the requirements for the degree of MSc.

in the

School of Mathematical and Computer Sciences

April 2025

Declaration of Authorship

I, Guilhem VINET, declare that this thesis titled, 'Detecting Fake Accounts on LinkedIn' and the work presented in it is my own. I confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Signed:
GV

Date:
10/04/2025

Abstract

LinkedIn is a Microsoft-owned social network that allows users to maintain and make contact with professional relationships.

This social network has been a growing platform with more than 1 billion users in more than 200 countries in 2025 and has become almost indispensable in many working industries.

Even if LinkedIn is a professional social network, the platform is targeted like the others (Twitter, Instagram, Facebook) by scams and phishing campaigns that overflow the platform. LinkedIn has tried to detect fake accounts with an AI-generated content detector and a feature that warns the user if the private message they receive seems suspicious.

As of 2025, few studies have been done in the case of LinkedIn; one study focuses on detecting the registration of clusters of fake accounts, and another one used data mining. The study aims to improve fake account detection on the platform LinkedIn by focusing on single account detection (when the account is already created). This work brings a new approach on LinkedIn with the usage of random forest and deep learning. In addition, we analyse additional features (Account Linguistic, AI-Generated Profile Pictures) to potentially improve the detection.

The outcome of this work is a reliable system that detects whether the information provided about the account corresponds to a genuine account or not.

Contents

List of Figures

Chapter 1

Introduction

1.1 Context

Social media shapes our lives both in our private and professional lives. These platforms are omnipresent in many aspects of our lives, and they should be safe and healthy for the users. Unfortunately, the most popular platforms (Twitter, Instagram, LinkedIn) are overflowing with bots and fake accounts.

The reported scams performed by LinkedIn LinkedIn [nd] show that the most common scams are Inheritance or advanced fee fraud scams, Job scams, technical support scams, and dating and romance scams.

Fake accounts make legitimate users vulnerable to scams, spam, and phishing NordLayer [2024]. 63% of scams in the United Kingdom are fake job offers; this is the most common scam on the platform in the United Kingdom. This is a serious issue when considering the amount of personal information that a candidate must send to get a job.

1.2 Motivations

This issue could lead to identity theft, which could be a disaster for the person whose identity is stolen Bharne and Bhaladhare [2022]. A research called the "Online recruitment services: another playground for fraudsters" Isaiah et al. [2022] made by Emil Eifrem, CEO of Neo Technology and co-founder of the Neo4j project, highlights the problem. This article concludes that information stolen from the fake job offer could be sold to legitimate third parties (cold-callers, marketers, political campaign operators) but also to malicious ones. They especially target the collection of sensitive information

like social security numbers, ID numbers, or bank account details to perform financial fraud.

In another way, fake accounts apply to legitimate job offers, this time to perform malicious activities. This happened in the United States of America. Indeed, North Korea targeted USA IT companies by applying to remote job offers. One North Korean candidate was using a fake LinkedIn account to get remote employment opportunities Group [2025]. In addition, in late 2024, Google's Threat Intelligence Group reported an evolution of these scams on European companies Greig [2025].

This finding changed the problem from a local threat in the US to a wide one that could inspire other malicious actors to perform the same kind of scams.

This research is also led by a personal motivation. I have been targeted by a fake internship offer in 2023 LinkedIn [2022]. The scam was advanced; the account that published it had a lot of followers, and the offer was put forward by Google Jobs and LinkedIn. I was already aware at this time of the potential of a scam online, so I checked the other job offers they had, and they got hundreds of copy-paste job offers; they were all the same, which is very suspicious.

1.3 Objective

The objective of this project is to improve the reliability of fake account detection on LinkedIn.

Previous research has been made on the case of LinkedIn, but they admit that their results are based on limited and static profile data and that it could be improved by using a larger dataset Vidros et al. [2016].

This work is different by the type of data analysed and the techniques used to do it. Instead of a limited number of labeled fake accounts (the previous work carried a database of 70 accounts), we use a bigger dataset of genuine accounts.

The techniques used are also different; this report aims to use Random Forest and Natural Language Processing to analyze account content.

To go further, we will analyze the account content with an Artificial Intelligence detection model to check if the content is AI generated; this is classified as suspicious behavior (for example, the North Korean scams are using AI generated profile pictures Isaiah et al. [2022]).

In addition to the model, we can access governmental public APIs to check if the companies that the account works for exist.

By using these different techniques, we can score the account; the lower the score is, the less the probability that the account is genuine. The performance of the system will be tested by comparing its results with accounts that have been labeled as fake.

1.4 Impact

This could help to avoid LinkedIn users getting scammed from fake accounts and improve the trust on the platform. Also, this research could be used to develop other tools for other platforms.

Chapter 2

Literature Review

2.1 Definitions

2.1.1 Fake account definition

A fake account, or Sybil, is an account that is meant to not respect the rules of the platform where it is registered. It's made to hide the real identity of someone who wants to do forbidden actions on the platform. This work focuses on the detection of fake accounts only.

In our case with LinkedIn, fake accounts spread online scams to make money and try to trick users with private messages to get sensitive data from the user to commit frauds. Fake accounts also perform Job scams on the platform, and according to LinkedIn, this is mainly for financial interestsLinkedIn [nd].

2.1.2 Deep learning definition

As explained on the Website of IBM IBM [nda], deep learning is a subfield of machine learning that uses artificial neural networks (ANN) with multiple layers to model and process complex patterns in data. This ANN take has been inspired by the human brain, the networks are composed of interconnected nodes. This node forwards propagation and refines predictions as data moves through each layer. This system is composed of 3 layers. The input layers receive the data that will be processed by the nodes. The hidden layer is composed of different sub-layers; this is where most of the computation happens. The results are pushed in the output layer to get the final prediction or classification. The input layer receives raw data, and the output layer delivers the model's final prediction or classification. IBM [nda]

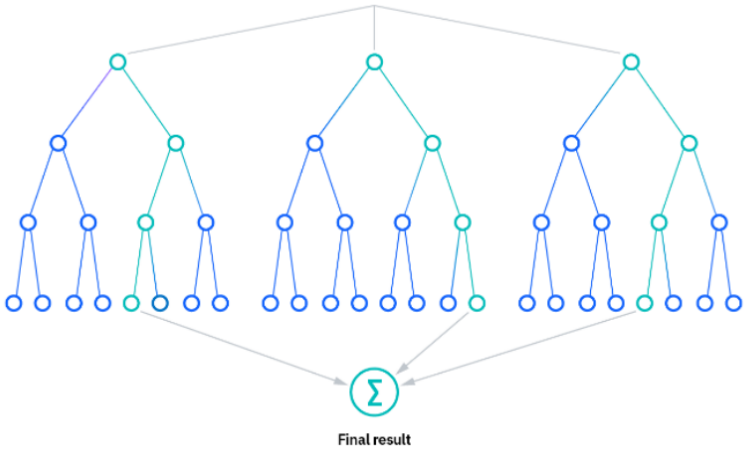

FIGURE 2.1: Figure 1: Random forest diagram IBM [ndb]

2.1.3 Random Forest definition

Random Forest is defined by Google developers as "an ensemble of decision trees, where each tree is trained using a specific form of random noise" Google [nd]. This type of decision tree ensemble is a popular technology to perform classification of data, in our case, to classify the account as genuine or otherwise fake. Random forests are composed of 3 main hyperparameters, which are the following: node size, the number of trees, and the number of features sampled. This parameter has to be set before the training to define how the random forest will classify or regress problems. Figure 1 shows an example of a random forest, it is composed of 3 decision trees. For the regression problem, each tree gives a number, and the final prediction is the average of all trees. For a classification problem, each tree votes for a class, and the final prediction is the majority vote.IBM [ndb]

2.2 Fake account detection background

Based on the work made in "A Review Article on Detection of Fake Profile on Social-Media"Roy and Chahar [2021] and on my findings, fake account detection has been studied for a decade, especially with the rise of Facebook in the 2010s. This subject has been explored in 2012 with the detection of Sybil attacks Wang et al. [2012]. In this research, the authors explore the detection of fake accounts by using a crowdsourced Sybil detection system for Online Social Networks. The crowdsourced method is not the most popular method used nowadays for fake account detection, but this research warns about the limitations of an automated fake account detection system. First, fake accounts do not struggle to connect with legitimate users, so based on the analysis of the

Category	Features from the Review	Features present on LinkedIn (public account)
Account	username profile photo biography	full name profile photo Description of the account
Social Metrics	following Count follower Count	connections
Location	location information on posts	Location (based on the information of the current job and past jobs)
Date	account creation date of the posts	Account creation Date of recent activity, posts, comments, likes

FIGURE 2.2: *Figure 2: Matching of features between FREQUENTLY USED FEATURES FOR FAKE ACCOUNT DETECTION*Roy and Chahar [2021] *and a public LinkedIn data account*

fake account network, a fake account will mainly have other fake accounts as contacts, which is irrelevant. This work also highlights the evidence of Syblis accounts that use advanced techniques to create realistic profiles, for example, copying data from genuine accounts to create fake ones. This information has to be taken into consideration for the development of my system to have the best accuracy possible and avoid false positives. On the other hand, the automatisation of the detection of fake accounts appears to be more popular, considering the numbers of publications made on this compared to the crowdfunding detection.

2.3 Commonly used account features

In "Fake Profile Detection on Social Networking Websites: A Comprehensive Review"Roy and Chahar [2021], they highlight the features of accounts that have been used in studies for fake account detection in online social networking (OSN). The authors based their work on multiple platforms, and it could be used for LinkedIn as the platforms share common features. The features are presented on a table named "FREQUENTLY USED FEATURES FOR FAKE ACCOUNT DETECTION"Roy and Chahar [2021]. I have selected the features from the table to do a comparison in Figure 2

This selection is useful to know what information could be relevant for an account base feature analysis because other studies prove their efficiency on their results.

There is a specific example of account features being used for fake account detection on LinkedIn in the paper "Identifying Fake Profiles in LinkedIn"Adikari and Dutta [2020], written by Shalinda Adikari and Kaushik Dutta.

Profile feature	Maximum value	Average value	Description
No_Languages	5	0.347	Number of languages spoken
Profile_Summary	1	0.52	Presence of profile summary
No_Edu_Qualification	7	1.467	Number of education qualifications attained
No_Connections	500	294.867	Number of connections to other profiles
No_Recommendation	37	2	Number of recommendations made
Web_Site_URL	1	0.28	Presence of a URL for personal web site
No_Skills	50	10.213	Number of skills and expertise listed
No_Professions	16	3.08	Number of past and present professions listed
Profile_Image	1	0.76	Presence of a profile image
No_Awards	10	0.56	Number of awards won
Interests	1	0.267	Presence of any type of interests
No_LinkedIn_Groups	51	8.907	Number of LinkedIn groups and associations added
No_Publications	16	0.613	Number of publications listed
No_Projects	7	0.24	Number of work projects listed
No_Certificates	9	0.267	Number of certificates held

Table 1: Details of the profile features

FIGURE 2.3: *Figure 3: Results of data collection from "Identifying Fake Profiles in LinkedIn"Adikari and Dutta [2020]*

They use a data mining approach for the detection. Using this approach, they got 87% accuracy on fake profile detection and 94% True Negative Rate. Even if they used the minimal amount of data necessary, they declared having results comparable to results obtained on a dataset with more accounts and more information on the accounts.

They collected data from public URLs and did not make any connection with the collected account to assure that the data are publicly available Figure 3 shows the results of the data collection, this represents the presence of the profile features and their average value on the 74 profiles collected for their dataset. This analysis could be done for the data classification on the datasets used in my work.

2.4 Artificial Neural Network (ANN)

LinkedIn made an official post where Jenelle Bray, Director of Engineering, Anti-Abuse AI, described the fake account detection used in 2018 (the time when the article was made) B. [2018].

They analyse the bulk registration of accounts, each new account creation is given a risk score by supervised machine learning models. If the fake account is not detected at registration, it will probably be detected by their activity-based models. They ensure that the majority of the fake accounts are

These features can be used for Machine Learning by using Deep learning and random forest as demonstrated in this research, "Fake Profile Detection Using Machine Learning Techniques"Chakraborty et al. [2022]. The author used Keras's' sequential model, which allows the creation of models layer by layer. By doing this, we can specify how many and what type of layers we need for the deep learning model. The authors put the

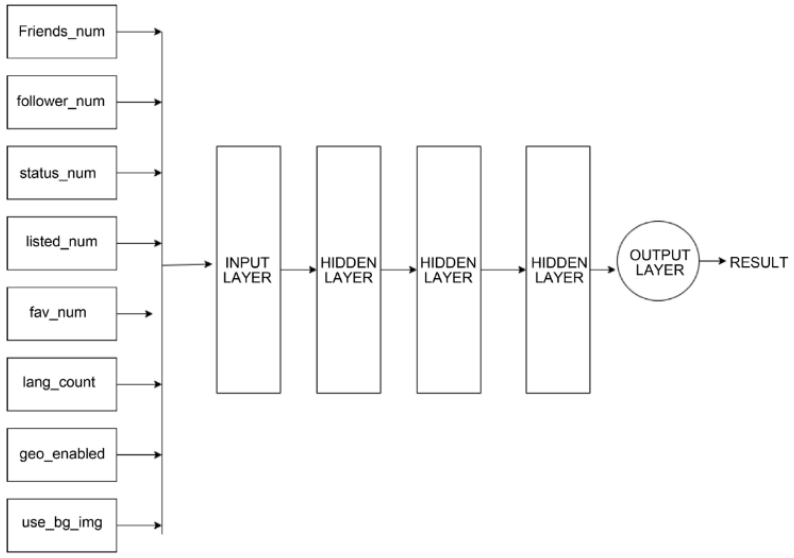

FIGURE 2.4: *Figure 4 Artificial Neural Network architecture used for fake account detection Chakraborty et al. [2022]*

account features in the input layer, and 3 hidden layers are used for the computation of the classification (see Figure 4). The input layer and hidden layers use activation functions to help the model learn complex patterns. In the output layer, the sigmoid activation function is used to convert the final output into a value between 0 and 1, which represents the model's confidence that a profile is either fake or real. The usage of Adam optimizer helps to train the model faster and more efficiently. With this system, the authors manage to get a starting accuracy of 0.97 and Peak accuracy of 0.98, which indicate an accurate detection of fake accounts in OSN by using deep learning. Using LinkedIn account features with deep learning could be a good way to detect fake accounts on the platform.

2.5 Random Forest

The research "Fake Profile Detection Using Machine Learning Techniques" Chakraborty et al. [2022] also brings results using random forest, a popular model for classifying data such as fake accounts (either the account is classy as genuine or fake).

The author declares having an accuracy of around 0.99 for decision trees and random forests.

Another source, titled "Fake Profile Identification using Machine Learning" Reddy [2019] proposes a detailed framework to perform fake account detection using random forest. The framework is composed of 6 parts.

S. No	Attribute	Description
1	Profile ID	The Profile ID of the account holder
2	Profile Name	The name of the account holder
3	Status Count	The number of tweets made by the account
4	Followers Count	The number of followers for the account
5	Friends Count	The number of friends for the account
6	Location	The location of the account holder
7	Created Date	The date the account was created
8	Share Count	The number of shares done by account holder
9	Gender	The gender of the account holder
10	Language Code	The language of the account holder

FIGURE 2.5: *Figure 5: Attributes selected by Mahatma Gandhi in his research "Fake Profile Identification using Machine Learning" Reddy [2019]*

First, the profiles are selected as inputs of the random forest, 80% of the dataset is used to prepare a training dataset, and the rest 20% is used for test dataset, both the dataset are composed of fake and genuine accounts. Datasets contain various attributes, so the author considered specific attributes (Figure 2.5).

This selection shares common account features (Created date, location, followers count, Profile name) with the table made in "Fake Profile Detection on Social Networking Websites: A Comprehensive Review"Roy and Chahar [2021].

This demonstrates the possibility of using random forest with LinkedIn account features(Figure 2.2) to detect sibyls. Once the selection is made, the attributes are passed into a trained classifier, the classifier gains accuracy on each iteration. Indeed, after this step, the profile is determined as fake or genuine, and this result is used as feedback to improve the classifier. This process is repeated, allowing the classifier to become more accurate over time.

The results are computed with a confusion matrix and a Receiver Operating Characteristic (ROC). By using this random forest framework, the author manages to identify fake profiles with an accuracy of around 95%. This should be the result that my research could aim for.

Other works have been applying random forest to detect fake accounts in OSN, such as Stefano Crescia, Roberto Di Pietrob, Marinella Petrocchia, Angelo Spognardia, and Maurizio Tesconia did in "Fame for sale: efficient detection of fake Twitter followers".Cresci et al. [2015]. To implement the random forest, they used the Weka framework, which is an open-source software. It's designed for machine learning, data mining, and data analysis.

This research carries out the detection of scam accounts and fake followers. This work got results with external datasets, which is very interesting for my work.

Results on external datasets 2.6:

1. On a test set of randomly sampled accounts:
   - Accuracy: 0.975
   - Precision: 0.982
2. Results on randomly sampled Obama followers:
   - Accuracy: 0.929
   - Precision: 0.889

Accuracy is defined as the proportion of predicted true results (both true positives and true negatives). In this context of fake follower detection, accuracy represents how often the classifier is correct overall. Precision is the prediction of positive cases that are indeed real positives. In the context of fake follower detection, precision is the percentage of accounts flagged as fake that are fake. The accuracy and precision of the random forest will be required to evaluate the usage of Sybils on LinkedIn.

In this research, they outline the advantages of using random forest compared to other classifiers. Random forest consistently outperforms other classifiers in most metrics (2.6). It also works well with both comprehensive and reduced feature sets.

Their research also revealed that for the fake Twitter follower (figure 2.7), certain features were particularly important for the classifier. Friends/(followersˆ2) ratio is the most influential feature according to the global sensitivity analysis.

This conclusion shows the importance of features for a random forest classifier; a similar analysis could be done in the case of a fake account on LinkedIn to check if any features stand out from the others.

Both of these studies Reddy [2019], Cresci et al. [2015] suggest the analysis of the content of the tweets to improve the accuracy of the detection; in our case for LinkedIn it will be the content of the posts and comments.

2.6 Content Analysis

By adding content analysis, we could expect to get better accuracy on the detection of fake accounts on LinkedIn. Analyzing the content of the posts and the comments made

	algorithm	evaluation metrics					
		accuracy	*precision*	*recall*	*F-M.*	*MCC*	*AUC*
Class A validation on a test set of random sampled accounts							
RF	Random Forest	**0.975**	**0.982**	**0.975**	**0.979**	**0.949**	**0.989**
D	Decorate	0.904	0.894	0.948	0.920	0.802	0.975
J48	Decision Tree	0.904	0.898	0.942	0.920	0.801	0.962
AB	Adaptive Boosting	0.767	0.737	0.936	0.825	0.526	0.959
BN	Bayesian Network	0.891	0.876	0.947	0.910	0.776	0.961
kNN	k-Nearest Neighbors	0.946	0.962	0.944	0.953	0.889	0.969
LR	Logistic Regression	0.551	0.922	0.252	0.396	0.299	0.827
SVM	Support Vector Machine	0.955	**0.982**	0.941	0.961	0.910	0.958
Class A validation on a test set of random sampled Obama followers							
RF	Random Forest	**0.929**	**0.889**	**0.975**	**0.930**	**0.862**	**0.970**
D	Decorate	0.909	0.875	0.948	0.910	0.820	0.964
J48	Decision Tree	0.902	0.868	0.942	0.903	0.807	0.924
AB	Adaptive Boosting	0.862	0.810	0.936	0.868	0.733	0.949
BN	Bayesian Network	0.786	0.710	0.947	0.811	0.607	0.943
kNN	k-Nearest Neighbors	0.733	0.763	0.655	0.705	0.469	0.828
LR	Logistic Regression	0.615	0.784	0.290	0.423	0.278	0.794
SVM	Support Vector Machine	0.873	0.851	0.897	0.873	0.748	0.874

FIGURE 2.6: *Figure 6 Class A classifier validation on two different test sets from*Cresci et al. [2015]

rank	feature	proposed in	normalized score
1	friends/(followers^2) ratio	[8]	1.000
2	age	[2, 28, 4, 1]	0.919
3	number of tweets	[1, 4, 8, 12, 14]	0.816
4	profile has name	[12]	0.782
5	number of friends	[32, 8, 14, 4]	0.781
6	has URL in profile	[12]	0.768
7	following rate	[2]	0.765
8	default image after 2 months	[14]	0.755
9	belongs to a list	[12]	0.752
10	profile has image	[12]	0.751
11	friends/followers $\geq$ 50	[14]	0.736
12	*bot* in biography	[11]	0.734
13	duplicate profile pictures	[11]	0.731
14	2 $\times$ followers $\geq$ friends	[11]	0.721
15	friends/followers $\simeq$ 100	[11]	0.707
16	has address	[12]	0.677
17	no bio, no location, friends $\geq$ 100	[14]	0.664
18	has biography	[12]	0.602
19	number of followers	[12, 4, 3]	0.594

FIGURE 2.7: *Figure 7 features for the detection of fake Twitter followers Cresci et al. [2015]*

	Profile	Behavior	Language	Stop word	BoW
Precision	0.07	0.24	0.15	0.14	0.58
Recall	0.01	0.4	0.21	0.39	0.79
F1	0.02	0.3	0.17	0.21	0.66
AUC	0.74	0.76	0.85	0.86	0.98
Accuracy	0.98	0.97	0.97	0.96	0.99

FIGURE 2.8: *Figure 8: Average performance of the Logistic Regression classifiers using each category of features when conducting 10-fold cross validation.*Im et al. [2020]

by the account has been used to detect adversarial attacks on fake account detection using machine learning Kantartopoulos et al. [2020].

The authors mention the usage of the shortened URLs that mask potentially malicious content (like phishing links). This is done due to the size limit (280 characters) of the tweets. In our case, a LinkedIn post has a 3000-character limit, so the same could be done to make the phishing link fit inside the post.

However, shortened URLs are not by definition a malicious behavior, this could also be done by a genuine account.

The detection of shortened URLs will have a low weight for detection with deep learning and random forest.

One approach to analyse user content is to use Linguistic Pattern Recognition. Indeed, if the account is labeled as an English native speaker (like a United States of America citizen but the detected language is French, this could be suspicious.

A paper analyzes the Russian trolls on Twitter Im et al. [2020] and uses "Stop word usage features".

This allows us to find specific types of words and their frequency of usage. This technique can make the difference between a non-native English speaker even if he is fluent and can recover the originally spoken language.Pemistahl [nd]

They had a result of 0.96 accuracy by using the stop word method (2.8). This could be used to analyze the posts and the comments of an account and determine the language of the account. Another detection will be carried out for this dissertation by using the python library Langua-py, which allows you to get the language of a text. LanguaPy is a Python library that supports 75 languages. It stands out for its efficiency on short text like tweets; this is fit for the detection of the language of LinkedIn posts and comments. The result of this analysis (if the language matches or not) is added to the features account to detect the fake account.

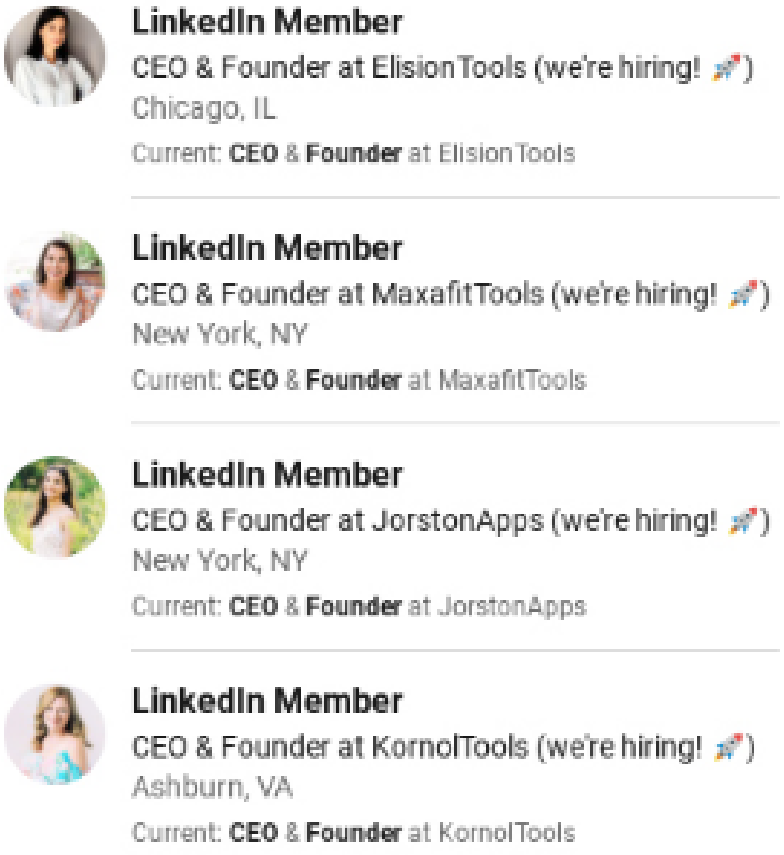

FIGURE 2.9: Figure 9 Duplicate profile generated by artificial intelligence Isaiah et al. [2022]

Artificial Intelligence content generation Another way to analyze LinkedIn's content is to detect if the content is AI-generated. Indeed, the most common feature to detect is that the profile picture is IA-generated. Users of the platform also interact with each other in real life, which means the usage of pseudonyms or hiding their real identity is not the popular usage of LinkedIn compared to other OSNs. So, most of the users are using real profile pictures of themselves.

A previous research has been made to warn on the LLM-generated LinkedIn Profiles. Ayoobi et al. [2023] . They highlight the fact that Large Language Models (LLMs) present a growing challenge for fake account detection. Indeed, these models can be used to: Generate content for multiple LinkedIn profile sections (Figure 2.9) Use a profile picture generated by artificial intelligence Reference personal information of target profiles to appear more legitimate Clone a genuine account

The detection of profile pictures made by artificial intelligence has been explored by Kai-Cheng Yang, Danishjeet Singh, and Filippo Menczer in the paper "Characteristics and prevalence of fake social media profiles with AI-generated faces" Yang et al. [2024]. They

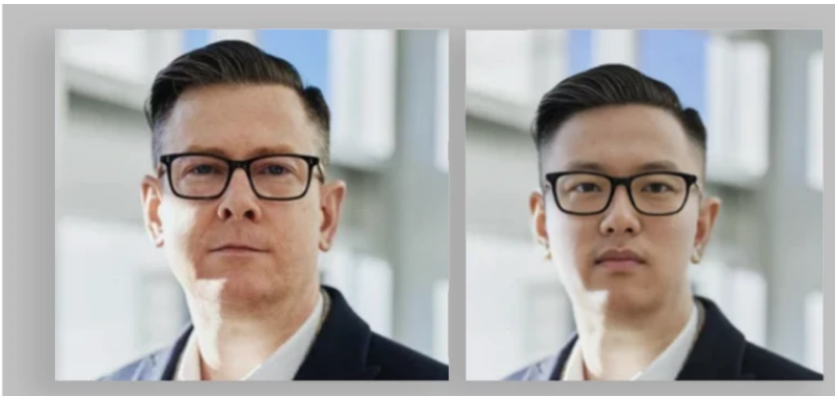

FIGURE 2.10: Figure 10 The original one (right) and the AI-generated one (left) Tidy [2023]

focus their work on identifying fake accounts and bots that use Generative adversarial networks(GAN) to generate faces; to do that, they mainly analyze profile pictures. Their model is published online Yang et al. [2024], so we could test it on the LinkedIn accounts. This model will be used to test if the profile picture of the LinkedIn account is generated by IA. The results of the model will add features to the analysis of the random forest and deep learning.

To generate the profile picture, GANs are being used with a real picture found on public image banks. This has been used by North Korean fake employees to get hired into American companies 2.10.

Chapter 3

Requirements

Task	Functional Requirement	MoSCoW	Priority	Description	Evaluation Metric / Notes
0	LinkedIn account features importance analysis on the dataset	M	1	Find the repartition of the features on the dataset	Get the importance of each feature of the account
1	Artificial Neural Networks (ANN)	M	1	Use deep learning to classify accounts	Compare with Random Forest using Accuracy, Precision, Recall, F1-score Confusion matrix and ROC curve analysis
2	Random Forest Classifier (RF)	M	1	Use random forest to classify accounts	Compare to ANN Accuracy, Precision, Recall, F1-score A confusion matrix and ROC curve analysis
3	System for features outside of the dataset	M	2	Create a system to add the ANN and the RF classifier data from other analyses (linguistic, IA-generated content, company verification, short URLs)	All additional features are usable for the ANN and the RF classifier
4	Analyze Account Linguistic	S	2	Analyze the language of the post and if the user is native in the language	Accuracy of language detection
5	Detect AI-Generated Profile Pictures	S	2	Analyze the profile picture to detect usage of IA	Accurate flagging of known AI-generated images
6	Analyze posts that contain shorten URLs	C	3	Flag posts that have shorten URLs in it	Accurate flagging of known post with shorten URLs
8	Company Verification via governamental REST API	C	4	Get location information from the jobs details and check on the APIs if the company is from the same country	Rate of correct company existence checks
9	A GUI to use the system	C	4	Allow the use of a graphical user interface to have a user-friendly system	test coverage (every test should pass)

FIGURE 3.1: Figure 11 Table of the functional requirements and their evaluations.

Task	Non-functional Requirement	MoSCoW	Priority	Description
0	Accuracy of at least 95% and precision of at least 90% for the Random Forest	M	1	What value is required to get a successful classification
1	Compare the Accuracy, Precision, Recall, and F1-score of the Deep learning model	M	1	Analyze the results between Random Forest and Deep Learning
2	Use public dataset and publicly available data	M	1	Avoid privacy and ethical issues

FIGURE 3.2: Figure 12 Table of the non-functional requirements and their evaluations.

Chapter 4

Professional, Legal, Ethical, and Social Issues

4.1 Professional issues

Developing a system for detecting fake accounts on LinkedIn involved professional issues. Indeed, ensure that every dataset used is unbiased and suitable for the analysis. In addition, the analysis process has to involve data validation techniques to ensure reliable results. These are done by using proven and rigorous testing techniques to ensure that the analysis is done professionally. Finally, the documentation of every technical point has to be provided to assure the reliability of the technology and their usages.

4.2 Legal issues

This work also involves legal issues; all the data has to be publicly available to use data without explicit permission (using only public data of a LinkedIn account). Also, data collection should comply with LinkedIn's terms of service. Data should be protected by implementing data anonymization.

4.3 Ethical

In addition, by developing this system, ethical issues are raised. The system has to be transparent by providing details on how data are used and with what techniques the classification is done. Any algorithmic bias and limitation has to be raised to ensure ethical work.

4.4 Social issues

Finally, if this system is adapted into a real-case scenario for a platform (LinkedIn or another one), it should not be used with cultural or ethnic bias (using a known country for scamming or botting as a reliable feature)

Chapter 5

Project Plan

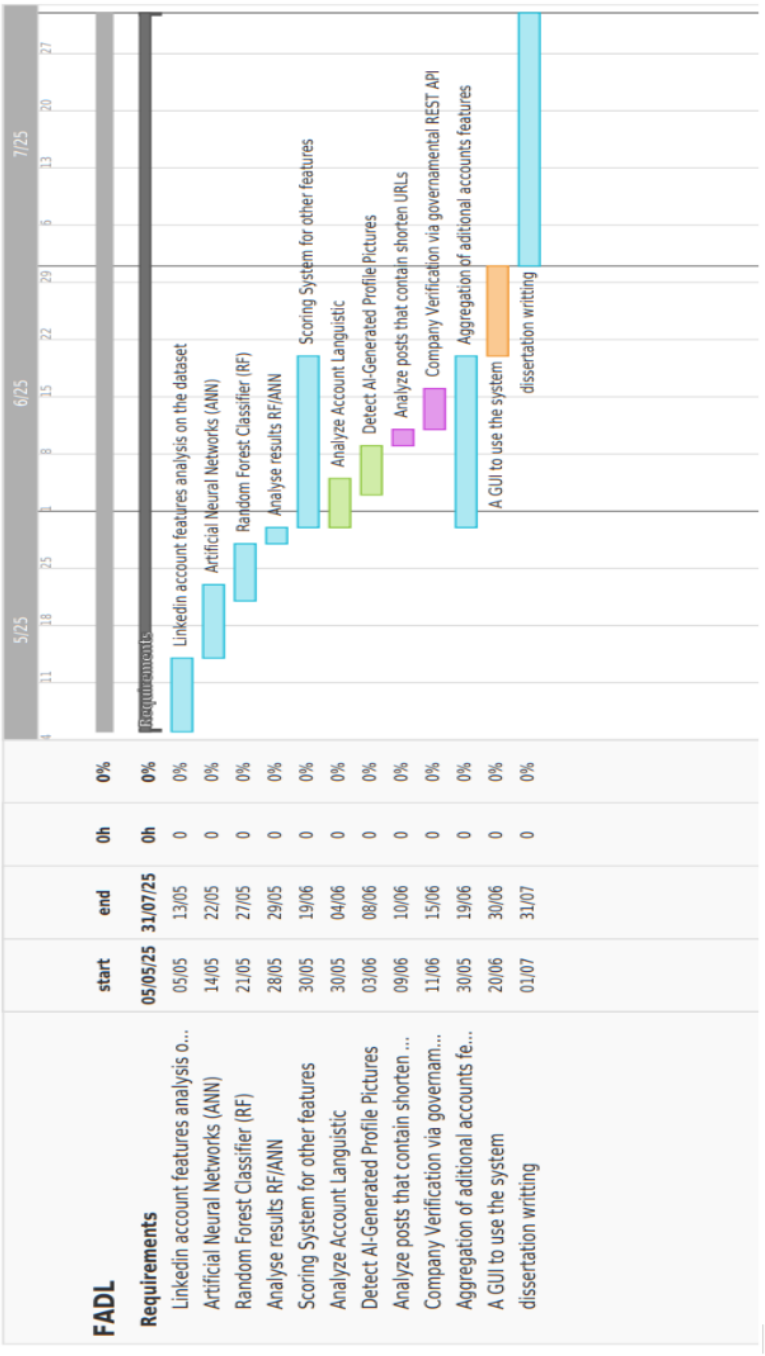

FIGURE 5.1: GANT chart

Chapter 6

Risk Analysis

6.1 Data risks

- Data quality could lead to bad performances for the classification model.
- Not enough data for the training of the deep learning classification.
- Usage of nonethical approval data.
- Dataset poisoning by malicious actors.

6.2 Classification model risks

- The training of the random forest could take more time than expected.
- The classification method used for Twitter and Facebook could not be efficient

for LinkedIn accounts.

The evolution of the methods to create fake accounts could make this work obsolete;

6.3 Additional features risks

- The experiments with the implementation of additional account features. (linguistic, IA-generated content, company verification, short URLs) could not do any improvement to the detection.
- The company verification could not be reliable over time due to the availability of the APIs and their limitations.

- IA-generated content will change over the evolution of IA, so the scalability of this feature could be compromised.

Chapter 7

Conclusion

7.1 Conclusion

In summary, fake account detection is a complex field that involves many different approaches (crowdsourced Sybil detection, data mining, artificial neural network, supervised Learning). Previous work shows that detection is possible on LinkedIn by doing data mining and detecting clusters of fake accounts. The main classification methods (deep learning and random forest) chosen rely on previous work made on Sybil detection on other OSN (Twitter, Facebook, Instagram). This relies on detecting the account based on the features of the account : Profile Name, location, current company, position. This work will carry a comparison between these two techniques on the case of LinkedIn.

As far as I know, no Sybil detection on LinkedIn has been done by analysing the content of the account. (Analyze Account linguistics, detect AI-generated profile Pictures, analyze posts that contain shortened URLs, and Company Verification via government REST API). These ideas came from reflection and discussion with my supervisor and future work proposal in papers about fake account detection in general. To prove their potential efficiency, we review papers that have a similar approach on other OSNs. By using these additional features, I expect more accurate results. This will involve the creation of a system to add the additional features and original accounts features to use them together in the models classifier.

However, this work has limitations, the main one being the data. The performance of the classification between genuine and fake will be determined by the callifier but also by the data used. So, the quality of the data will impact the results.

The usage of Analyze Account Linguistic, Detect AI-Generated Profile Pictures, Analyze posts that contain shortened URLs, and Company Verification via government REST API has not been proven yet for Sybil detection on LinkedIn.

Finally, this work is focused on the detection of fake accounts only, that does not include clone accounts, bots and spams. Other studies could carry this missing detection in the future.

Appendix A

Appendix

Dataset sources:

https://www.kaggle.com/code/rajatraj0502/linkedin-professional-profiles-dataset

Bibliography

Adikari, S. and Dutta, K. (2020). Identifying fake profiles in linkedin. *arXiv preprint arXiv:2006.01381*.

Ayoobi, N., Shahriar, S., and Mukherjee, A. (2023). The looming threat of fake and llm-generated linkedin profiles: Challenges and opportunities for detection and prevention. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*, pages 1–10.

B., J. (2018). Automated fake account detection at linkedin. https://www.linkedin.com/blog/engineering/trust-and-safety/automated-fake-account-detection-at-linkedin.

Bharne, S. and Bhaladhare, P. (2022). Comprehensive analysis of online social network frauds. In *International Conference on Advances in Data-driven Computing and Intelligent Systems*, pages 23–40. Springer Nature Singapore.

Chakraborty, P., Shazan, M., Nahid, M., Ahmed, M., and Talukder, P. (2022). Fake profile detection using machine learning techniques. *Journal of Computer and Communications*, 10(10):74–87.

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., and Tesconi, M. (2015). Fame for sale: Efficient detection of fake twitter followers. *Decision Support Systems*, 80:56–71.

Google (n.d.). Random forest - google ai. https://developers.google.com/machine-learning/decision-forests/random-forests#:~:text=A%20random%20forest%20(RF)%20is,form%20of%20decision%20tree%20ensemble.

Greig, J. (2025). Europe: Europe north korea fake employees. https://therecord.media/north-korean-it-worker-scam-spreads-to-europe.

Group, I. (2025). North korea fake employees. https://www.recordedfuture.com/research/inside-the-scam-north-koreas-it-worker-threat.

IBM (n.d.a). What is deep learning? https://www.ibm.com/think/topics/deep-learning.

IBM (n.d.b). What is random forest? https://www.ibm.com/topics/random-forest.

Im, J., Chandrasekharan, E., Sargent, J., Lighthammer, P., Denby, T., Bhargava, A., Hemphill, L., Jurgens, D., and Gilbert, E. (2020). Still out there: Modeling and identifying russian troll accounts on twitter. In *Proceedings of the 12th ACM conference on web Science*, pages 1–10.

Isaiah, J., Caitlin, M., and Amie, N. (2022). Rise of ai-generated, fake linkedin profiles raises social engineering challenges. https://www.kroll.com/en/insights/publications/cyber/rise-of-ai-generated-fake-linkedin-profiles-social-engineering-challenges.

Kantartopoulos, P., Pitropakis, N., Mylonas, A., and Kylilis, N. (2020). Exploring adversarial attacks and defences for fake twitter account detection. *Technologies*, 8(4):64.

LinkedIn (2022). Personal post on linkedin. https://www.linkedin.com/posts/guilhem-vinet-661572190_rjobboardsearch-on-reddit-talentkompass-activity-7112446339409408000-QxSR?utm_source=share&utm_medium=member_desktop&rcm=ACoAACzvBCkBIsax-daA-qt5NxYAGpa-Ot0ittY.

LinkedIn (n.d.). Recognize and report scams. https://www.linkedin.com/help/linkedin/answer/a1336387.

NordLayer (2024). How to identify and avoid linkedin scams. https://nordlayer.com/blog/linkedin-scams/.

Pemistahl (n.d.). Lingua.py. https://github.com/pemistahl/lingua-py.

Reddy, S. (2019). Fake profile identification using machine learning. *International Research Journal of Engineering and Technology (IRJET)*, 6(12):1145–1150.

Roy, P. and Chahar, S. (2021). Fake profile detection on social networking websites: a comprehensive review. *IEEE Transactions on Artificial Intelligence*, 1(3):271–285.

Tidy, J. (2023). Firm hacked after accidentally hiring north korean cyber criminal. https://www.bbc.co.uk/news/articles/ce8vedz4yk7o.

Vidros, S., Kolias, C., and Kambourakis, G. (2016). Online recruitment services: another playground for fraudsters. *Computer Fraud & Security*, (3):8–13.

Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M., Zheng, H., and Zhao, B. Y. (2012). Social turing tests: Crowdsourcing sybil detection. *arXiv*.

Yang, K., Singh, D., and Menczer, F. (2024). Characteristics and prevalence of fake social media profiles with ai-generated faces. *arXiv*. [online].