## ***Disclaimer:*** *The following document is intended to provide general guidance on the use of this template. Please refer to the template for the specific format and content requirements. You may also add additional information to the template as needed for your specific project.*

## **Group 1-5 Mid-Semester Project Evaluation Report**

**Project Title**: Detection Process of Suspicious Activities on Social Media Using Data Mining and Machine Learning.

## 

## List of members

**Group Leader:** [Krupaben Kothadia](mailto:kkothadi@asu.edu)

**Group Deputy Leader:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu)

**Group Members:** [Krupaben Kothadia](mailto:kkothadi@asu.edu)

* [Gautham Vijayaraj](mailto:gvijaya6@asu.edu)
* [Avani Mundra](mailto:amudra@asu.edu)
* [Rahul Nayak](mailto:rrnayak@asu.edu)
* [Justin Young](mailto:jtyoun15@asu.edu)
* [Anuranjan Dubey](mailto:adubey37@asu.edu)
* [Sangeeth Santhosh](mailto:ssantho9@asu.edu)
* [Yeshwanth Reddy Chennur](mailto:ychennur@asu.edu)

## Project Progress

* The team has completed the following tasks:

[Krupaben Kothadia](mailto:kkothadi@asu.edu)

* + - [An Integrated approach for Malicious Tweets detection using NLP](https://drive.google.com/file/d/1z9qYcbTsDz-IZOJNe5SgFdwLh9_qKNxi/view?usp=drive_link)**:** The paper shows how to analyze tweets better while keeping people's privacy safe. They used Twitter's tools to get some tweets that were already marked as spam or not spam. Advanced NLP techniques were applied to identify semantic meanings between words. The outcome is a language model with text-URL divergence classifying tweets into spam and non-spam. In-depth report link: [Krupaben\_Kothadia\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1Bo-t0owzhAoLDXxKKcmmtBKy1A3-otce/edit?usp=drive_link&ouid=107121383631492907164&rtpof=true&sd=true)
    - [Intelligent Analysis of Arabic Tweets for Detection of Suspicious Messages](https://drive.google.com/file/d/1AA1mL9x8Nbc39Pn1_JSlQLd4bKAZ32Vd/view?usp=drive_link)**:** The paper discusses challenges in analyzing Arabic tweets for suspicious activity detection in the context of the project. The findings reveal that Arabic tweets, particularly those without diacritics, present difficulties in meaning interpretation, and stemming and lemmatization were necessary for classification. The dataset consisted of 1555 tweets, with 826 categorized as suspicious and 729 as not suspicious, highlighting the importance of data preprocessing in the detection process. In-depth report link: [Krupaben\_Kothadia\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1PsIaYpLgsXiAWfzaVHTFckO0JiOgwIPs/edit?usp=drive_link&ouid=107121383631492907164&rtpof=true&sd=true)
    - [Mining user Message Pattern for Suspicious Behavior on Terrorism using NLP in Social Networks with Single Sign-On](https://drive.google.com/file/d/1STLuTC9mJx8LF8fihELth3KnG4n-0rLv/view?usp=drive_link): The paper discussed in the context of the project utilizes NLP and image analysis techniques to proactively identify terrorism-related threats within Social Networks, focusing on safeguarding lives and property. It successfully combines text and image data for suspicious activity detection, primarily drawing data from Gmail and Twitter through Single Sign-On authentication. The NLP module, powered by Apache OpenNLP, enhances the analysis of input data, enabling the extraction of valuable data patterns to aid in the detection process. In-depth report link: [Krupaben\_Kothadia\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1uPu-vDVb5l1wdpXaiZBJy_72-zk4GdHK/edit?usp=drive_link&ouid=107121383631492907164&rtpof=true&sd=true)

*(add information about reference papers or articles reviewed by the team members) (add info about the paper with respect to the project and the findings or important information from the same [please provide the drive link of the paper])*

[Gautham Vijayaraj](mailto:gvijaya6@asu.edu)

* + - [An Overview of Privacy Preserving Data Mining](https://drive.google.com/file/d/1XipNBU2_EKn2yklo7fJbQaWyXa1nXMy4/view?usp=drive_link)**:** The main focus area of this paper is to find different methods of maintaining privacy while performing data mining. The study reveals that the research of privacy protection methods are summarized by either performing privacy preserving data mining algorithms or using privacy protection technologies. Privacy protection methods such as distortion, encryption, privacy and anonymity are discussed. In-depth report link: [Gautham\_Vijayaraj\_Individual\_In-depth\_Report](https://docs.google.com/document/d/1nz6jbcDmSwb_I6wlzNtFrsAWNSx-vtjDxdMcVeeP8L8/edit?usp=drive_link)
    - [Privacy-Preserving Data Mining: Why, How, and When](https://drive.google.com/file/d/1JF6NG6eQJ1_Y9zLhEyODuRKGJUtQuPMf/view?usp=drive_link)**:** The main focus area of this paper is to find how technology from the security community can change data mining for the better, providing all its benefits while still maintaining privacy. If the SMC approach is used, the process of data mining doesn’t cause, or even increase the opportunity for, breaches of privacy because only selected parties have access to the data. However, multi-level security and accuracy in terms of efficiency needs improvement. In-depth report link: [Gautham\_Vijayaraj\_Individual\_In-depth\_Report](https://docs.google.com/document/d/1Kntk1wZENQjdTkGypBMYCTaT3hQYSbiYDZNCdxDpCyc/edit?usp=drive_link)
    - [Sentiments Analysis Of Twitter Data Using Data Mining](https://drive.google.com/file/d/1SGNHuQyCJ--drDbYgOgs12MNIApcoHFT/view?usp=drive_link)**:** The main focus area of this paper is to find an approach to analyze the sentiments of users using data mining classifiers.The polarity value of the tweets in Twitter was taken into consideration for finding suspicious activities. Experimental results obtained demonstrate that k-nearest neighbor classifier gives very high predictive accuracy with anaccuracy of 99.6456%. In-depth report link: [Gautham\_Vijayaraj\_Individual\_In-depth\_Report](https://docs.google.com/document/d/1fG4LzSGURuVtwNBOD_D3LfmGkpmiGffUHLq4mGQ0P58/edit?usp=drive_link)
    - [Comprehensive Survey on Big Data Privacy Protection](https://drive.google.com/file/d/1lnxy_kihxdIVJHpah0283JycADq59Jwx/view?usp=drive_link)**:** The main focus area of this paper is to elaborate on the existing challenges and unresolved issues in privacy-preserving data mining (PPDM). The current review study aims to utilize the tasks of data mining operations without risking the security of individuals’ sensitive information, particularly at the record level. The techniques involved in preserving privacy of data in data providers and data collectors are explored.The other approaches to privacy preserving data mining techniques include data modification approaches like k-anonymity approach, l-diversity approach, and t-closeness approach. In-depth report link: [Gautham\_Vijayaraj\_Individual\_In-depth\_Report](https://docs.google.com/document/d/1Jk-JD1tfp1L-R5b2Bo_3xJNAwFkwljtSn30yPU3LmUE/edit?usp=drive_link)

[Avani Mundra](mailto:amudra@asu.edu)

* + - [Cybercrime Profiling: Text mining techniques to detect and predict criminal activities in microblog posts](https://drive.google.com/file/d/15P2Qk-7sbmul830CMnhQu9Z2BBlGHgv-/view?usp=drive_link)**:** The main focus area of this paper is to overcome the challenges of data sparseness and reduce semantic gap in text mining techniques to detect suspicious behavior in Twitter posts. It extends a previous solution for detecting suspicious profiles, emphasizing disambiguation and semantic enrichment, with plans to further improve execution time and precision rates in future work. An example containing two suspicious words is used to prove the above approach. The similarity function returns 0, thereby declaring the content suspicious. In-depth report link: [Avani\_Mundra\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1LPoCyQ8MVEIZnsYgoUjSDU6hDPZErFUb/edit?usp=sharing&ouid=112867995748295868322&rtpof=true&sd=true)
    - [An approach to detect abusive content incorporating Word2Vec and Multilayer Perceptron](https://drive.google.com/file/d/109jQngdH9mkZbf2HCHX5phKmgxxt-Fy8/view?usp=drive_link)**:** The main focus area of this paper is to detect suspicious activities and hateful content in social media using deep learning techniques of Word2Vec and Multilayer Perceptron Neural Network. The compositional vector model computes tweet vectors from word vectors using the Bi-function, capturing bi-gram information and word interactions with hyperbolic tangent functions.The Multilayer Perceptron classifier is employed to classify tweet vectors obtained from the compositional vector model.The model achieved 86% accuracy and 0.84 weighted F1 score. In-depth report link: [Avani\_Mundra\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1C3yYKzLyBr2hDLX69AEa7T-Pg2kC0LYn/edit?usp=sharing&ouid=112867995748295868322&rtpof=true&sd=true)
    - [A Framework to Detect and Prevent Cyberbullying from Social Media by Exploring Machine Learning Algorithms](https://drive.google.com/file/d/1eTMuAMKP-NJ3s6PQh2fSY7WIHpw-w1Gq/view?usp=drive_link)**:** The main focus area of this paper is to tackle the escalating issue of cyberbullying within the context of social media, with a primary focus on Twitter.Two machine learning models are developed: one for detecting the sentiment of cyberbullying instances and the other for categorizing cyberbullying into different types.The paper introduces a novel conceptual framework, comprising three integral modules:User Interaction, Analysis, and Decision Making. Random Forest algorithm with TF IDF Embedding achieved impressive results with an F1 score of 80.8% for identifying cyberbullying and 58.4% for classifying the type of cyberbullying. In-depth report link:[Avani\_Mundra\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1FlVzEB8gLT0f0hTLmuGgjyHVz6Stfvj7/edit?usp=sharing&ouid=112867995748295868322&rtpof=true&sd=true)

[Justin Young](mailto:jtyoun15@asu.edu)

* + - [Comparative\_Analysis\_of\_Machine\_Learning\_Models\_in\_Computer\_Network\_Intrusion\_Detection.pdf](https://drive.google.com/file/d/136RWM4WEDzXD0z5dVBgdFyQoJZ_fdQ3n/view?usp=drive_link)**:** This paper focuses on performative comparison between various machine learning models in relation to network intrusion detection. A binary classification procedure is proposed by the study to determine how accurate each model is in classifying network activity as either malicious or not. The study concludes that the decision tree model performed the best among the others. In-depth report link: [Justin\_Young\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/12gjLQIFTVQhlbpW4wLfmzPYmFYMdt6jH/edit?usp=drive_link&ouid=117665874805656519311&rtpof=true&sd=true)
    - [Behavior-Based\_Malware\_Detection\_System\_Approach\_For\_Mobile\_Security\_Using\_Machine\_Learning.pdf](https://drive.google.com/file/d/1pSKALJaUpDPuxOn2EdftwNHr5kkSsHPg/view?usp=drive_link): This paper focuses on analyzing ensuring security on mobile devices using behavior based machine learning approaches to detect malicious content. This study provides a mobile framework which trains various models to be compared on the Py-Charm dataset. This paper concludes that the Naive Bayes model was the most accurate in detecting malicious content on mobile applications with an accuracy percentage of 97.37%. In-depth report link: [Justin\_Young\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/18Ie_EKDA87LVddDpUgR7epngBBpj6W65/edit?usp=drive_link&ouid=117665874805656519311&rtpof=true&sd=true)
    - [Cooperative\_Machine\_Learning\_Techniques\_for\_Cloud\_Intrusion\_Detection.pdf](https://drive.google.com/file/d/1_W5v281OJeIt1GDVTfCHtyv6cKP6tA4h/view?usp=drive_link): This paper focuses on intrusion detection specifically related to cloud based systems, proposing a new model called Secure Packet Classifier which aims to detect anomalies with collaborative filtering. To measure the performance of this model, a data set is partitioned and other machine learning models are trained on it for comparison. The results of this experiment concluded that the SPC model performed with ~20% better accuracy than traditional ML methods with an accuracy of 81%. In-depth report link: [Justin\_Young\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1ngfCZIp47B_1vlbMLniVQSB9svgv5RA7/edit?usp=drive_link&ouid=117665874805656519311&rtpof=true&sd=true)
    - [Machine\_Learning-based\_web\_security\_intrusion\_detection\_system.pdf](https://drive.google.com/file/d/1nvHttnxhIMSzOcZgFpUkXj1C3BKfgqEs/view?usp=drive_link): This paper focuses on comparing the performance of various machine learning models against attacks prominent in web security, SQL injection and XSS. This study analyzes the principles, classifiers and statistical data on these types of attacks to compare each model. The result of this study proves that while each model performed well against these attacks, the Adaboost model was most effective against SQL injection attacks, and the SVM model against XSS. This paper concludes that while these models overall performed well, they also explain their weaknesses. In-depth report link: [Justin\_Young\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1zGuy9dXoWNVJkjFfn_tm10JqR74BcCmc/edit?usp=drive_link&ouid=117665874805656519311&rtpof=true&sd=true)

[Anuranjan Dubey](mailto:adubey37@asu.edu)

* + - [Cyberthreat Detection from Twitter using Deep Neural Networks.pdf](https://drive.google.com/file/d/1NusmoTBnGaOwuXmbHVApEJaraEF95o0H/view?usp=drive_link): The main focus area of the research paper centers on employing deep neural networks, including CNN and BiLSTM, to classify security-related content on Twitter with remarkable precision. It streamlines threat intelligence gathering, automatically sifting through tweets, and identifying valuable information. The paper's efficacy is confirmed through simulations and real-world implementation, offering practical value to security analysts.  
      In-depth report link: [Anuranjan\_Dubey\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1NKiEEWwR1V9-8DPF39XKEpvrRfBJmBiA/edit?usp=drive_link&ouid=101870164252766530632&rtpof=true&sd=true)
    - [CyberTwitter- Using Twitter to generate alerts for Cybersecurity Threats and Vulnerabilities.pdf](https://drive.google.com/file/d/14LY8rxklXibD4hygSqNHTrHp7BaC_91m/view?usp=drive_link): The paper focuses on using the SVCE (Security Vulnerability Concept Extractor) to analyze Twitter data, extracting cybersecurity vulnerability-related entities. It employs SWRL for semantic meaning, RDF Triples for structured data extraction, and NER for entity classification. The CyberTwitter system generates personalized alerts based on real-time system profiles and semantic knowledge base data. The paper concludes by discussing improvements in tweet selection and word embedding models, with an acknowledgment of the absence of standard metrics to confirm its method's superiority. The research outcome demonstrates CyberTwitter's effectiveness in extracting valuable cybersecurity insights from Twitter during a ten-day experiment, affirming its potential for issuing meaningful threat alerts.  
      In-depth report link: [Anuranjan\_Dubey\_Individual\_In-depth\_Report](https://docs.google.com/document/d/1ZmY3NUKcXkcmkWu-1AncXaRacKjpBIbfnSPo0TnGBso/edit?usp=drive_link)
    - [Weakly Supervised Extraction of Computer Security Events from Twitter.pdf](https://drive.google.com/file/d/1FfclqLsECMhNk0dr1bzHV5BQhCBrWNvJ/view?usp=drive_link): The paper compared various algorithms, including One-Class SVMs and Expectation Regularization, with the latter displaying superior precision and recall for security content classification on Twitter. It employed historical seed instances for training, outperforming previous methods in terms of precision and recall. While specific accuracy percentages were not provided, the weakly supervised seed-based approach demonstrated substantial improvements, particularly for security-related events like DoS attacks and data breaches. The study underlined Twitter's value as a resource for security event information and the effectiveness of the proposed method.  
      In-depth report link: [Anuranjan\_Dubey\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1l3P5Ti-SS2k1O8JsMPobUNxNJhU897y9/edit?usp=drive_link&ouid=101870164252766530632&rtpof=true&sd=true)
    - [An effective security alert mechanism for real-time phishing tweet detection on Twitter.pdf](https://drive.google.com/file/d/1b6pghiZ9c4T-rlq8Gq_CX7Q2HIabxIC9/view?usp=drive_link): In this paper, the success of the security alert mechanism hinged on robust feature extraction. It employed a set of 22 diverse features, augmented by 7 more suggested by WEKA, bolstering its ability to distinguish phishing from legitimate tweets. The Random Forest machine learning technique was a pivotal choice, given its adaptability to Twitter's dynamic environment. Beyond simulations, real-world implementation proved the mechanism's efficiency in promptly identifying phishing URLs on Twitter. It achieved a remarkable 97.50% accuracy when evaluated with real datasets, ensuring rapid user alerts.  
      In-depth report link: [Anuranjan\_Dubey\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1-LaUI_L5yJk-uHkHieD2GH9BaNHmpD7F/edit?usp=drive_link&ouid=101870164252766530632&rtpof=true&sd=true)

[Sangeeth Santhosh](mailto:ssantho9@asu.edu)

* + - [Classifying Suspicious Content on Social Media Networks](https://drive.google.com/drive/u/2/folders/1Ke86Jww8XdcY1CXonsX9Lvy47MdYGCUu): In this paper, the authors propose a model for generation and classification of tweets into the categories of suspicious or non-suspicious using Sentiment Analysis and Machine Learning algorithms. This research paper does not compromise the integrity of honest users, while also being a source of great knowledge to social media users by making them aware of fake accounts and spam content.

In-depth report link: [Sangeeth\_Santhosh\_Individual\_In-depth\_Report](https://docs.google.com/document/d/1Eif5cCu7J-lSJmQg8JUva7IBHbHFUM5fCpB289qwmzY/edit?usp=drive_link)

* + - [A new approach for the detection and analysis of phishing in social media networks : the case of Twitter](https://drive.google.com/drive/u/2/folders/1Ke86Jww8XdcY1CXonsX9Lvy47MdYGCUu): The paper proposes an approach to detect and analyze phishing in Twitter. The proposed approach has three steps - Verification of URL occurrence in database of malicious URLs (Blacklisting), Classification of URLs using supervised machine learning, User Analysis Mechanism. Different machine learning algorithms are used to compare the results. Overall, this approach minimizes the number of false negatives and keeps the integrity of honest users in check.

In-depth report link:[Sangeeth\_Santhosh\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1B1hxOr_ErJt8EaPqAnqt_V5yt71pVEVB/edit?usp=drive_link&ouid=100662470547641371728&rtpof=true&sd=true)

* + - [Exploring and Detecting Opinion Spam on Social Media](https://drive.google.com/drive/u/2/folders/1Ke86Jww8XdcY1CXonsX9Lvy47MdYGCUu): This paper focuses on the detection and analysis of opinion spam on social media, while also proposing an approach to deal with such spam. The approach proposed has two steps - Empirical Quantitative Analysis and Context Based Classification. The f1 score is the main parameter used to compare the performance of these two steps. Thus, the research paper correctly labels users as spam or non-spam, while not incorrectly classifying honest users as spam.

In-depth report link: [Sangeeth Santhosh\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1iS3-fIrzOwuBhX32X8MVXcnS3Jkq8wkv/edit?usp=drive_link&ouid=100662470547641371728&rtpof=true&sd=true)

* + - [Spam Filtering of Bi-Lingual Tweets Using Machine Learning](https://drive.google.com/drive/u/2/folders/1Ke86Jww8XdcY1CXonsX9Lvy47MdYGCUu): In this paper, the spam filtering of bi-lingual tweets, mainly Roman Urdu tweets, is done using machine learning techniques. The approach proposed in the paper to effectively filter the tweets involves the following steps - Tweet Collection, Pre-processing, Data Preparation and Classification. The different machine learning algorithms come into picture during the classification process. Each of the algorithms are used for classification, their accuracy results are compared and the best approach chosen. Overall, the paper achieves successful classification of tweets by comparing the accuracy and ROC AUC parameters.

In-depth report link: [Sangeeth Santhosh\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1Ldh1ldPvBRYJMhr7a66R4gR6AKlbWJU3/edit?usp=drive_link&ouid=100662470547641371728&rtpof=true&sd=true)

[Yeshwanth Reddy Chennur](mailto:ychennur@asu.edu)

* [Towards automated real-time detection of misinformation on Twitter](https://drive.google.com/file/d/12Lm3n-1HkhbutIC-yg4k6kkTWQTXhiwn/view?usp=drive_link): The authors present a thorough method for detecting incorrect information in tweets that uses language and environmental factors. Their method is highly effective since it employs powerful machine learning algorithms for classification. This research greatly contributes to the subject of information verification and trustworthiness in online environments by addressing the current challenge of fighting disinformation on social media platforms. [Yeshwanth\_Reddy\_Chennur\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1cVgJIqDU6iaLSsUwmRFxPiTLrb41k1qK/edit)
* [A Trusted Social Network Using Hypothetical Mathematical Model and Decision- Based Scheme](https://drive.google.com/file/d/15-gauHeasPzFDcXhp6a1OyeXj3vuIdrT/view?usp=drive_link): Their solution is based on a hypothetical mathematical model and a decision-based strategy, which work together to boost user confidence by addressing privacy concerns and ensuring secure communication. The proposed model protects sensitive data effectively while striking a careful balance between privacy and engagement. This study is a significant step forward in the development of more secure and dependable social networking platforms, and it holds huge promise for the future of online social interactions.

[Yeshwanth\_Reddy\_Chennur\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/11vKsKw3MoRBUCHBcrFRM8hQpJ7R7Yqek/edit?usp=drive_link&ouid=100701464624984336471&rtpof=true&sd=true)

[Rahul Nayak](mailto:rrnayak@asu.edu)

* + - [Suspicious Behavior Detection Current Trends and Future Directions](https://drive.google.com/file/d/1DSvNUA1puHTXPRyASPjqKZtSWrpKDL1-/view?usp=drive_link)**:** This paper introduces about the different suspicious activities that go on social media. These types of suspicious activities on social media platforms can have detrimental effects, including the spread of misinformation, erosion of trust, and manipulation of user interactions. Detecting and mitigating these behaviors is crucial to maintain the integrity and security of online social environments.These are some of the suspicious behaviors that the paper discusses about Social Spam, Social Sybils, Astroturfing and Link Farming. [Rahul\_Nayak\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1s8XIZ5oJMw-bhkVqyn9PVmg4kDoLF9xE/edit?usp=drive_link&ouid=117613874970212534075&rtpof=true&sd=true)
    - [Malicious Profile Detection on Social Media: A Survey Paper](https://drive.google.com/file/d/1GSNFGH-WgLtx_cqDF3xERZ7DqdDa_zdf/view?usp=drive_link)**:**This research paper investigates the phenomenon of Malicious Profiles on social media, which constitutes a form of suspicious behavior within these platforms. The study includes a survey that reveals the wide range of suspicious activities associated with fake profiles, such as the existence of bot accounts, spam accounts distributing unwanted content, and duplicate accounts pretending to be real users. The paper presents an array of machine learning methods, including support vector machines, random forests, neural networks, and decision trees, that have been utilized to detect fake profiles on various social media platforms. [Rahul\_Nayak\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1QGkYKtKZymmKfcjz_5n3PdStGJ55IFFY/edit)
    - [Machine Learning Based Twitter Spam Account Detection A Review](https://drive.google.com/file/d/1Lr1zO-gHD-HFFDROSCib9VttNNWNZBF_/view?usp=drive_link):The paper discusses online social networks' growth and significance, highlighting their communication role. It acknowledges issues like spam, malware, and malicious activities on platforms such as Twitter. The paper aims to tackle Twitter's spam detection challenges with a machine learning framework comprising training and testing phases, employing metrics for performance evaluation.[Rahul\_Nayak\_Individual\_In-depth\_Report.docx](https://docs.google.com/document/d/1ZA5s9wc2Tp7KwbqorLPad3fPBkbkKb8M/edit#heading=h.gjdgxs)

We conducted zoom meetings every week, prepared our Work\_Distribution\_Draft so that team members can keep track of their tasks, and take meeting notes regularly. Every individual member gets an opportunity to prepare a weekly report and GANTT Chart, and visit the writing center on a rotational basis. We organized the Google drive every week by creating folders for every week’s name.

*[Add a paragraph for top 5 inferences from overall in-depth reports and mention all of us are working on core domains of our project, for more info about inferences refer to the particular paper mentioned above]*.

Individual Contribution

* Member Name: [Krupaben Kothadia](mailto:kkothadi@asu.edu)
  + I played a crucial role in several aspects of our project. I led the preparation of our Midterm Report, ensuring it was comprehensive and well-structured. Consistently, I evaluated, approved our weekly reports. I have prepared a weekly report for one time providing a clear overview of our team's progress. I got an opportunity to create and maintain the GANTT chart twice, helping us manage project timelines effectively. Additionally, I drafted a detailed work distribution plan, facilitating task allocation among our team members. I have written a summary report on 3 reference papers.
  + I also scheduled and facilitated our regular weekly meetings, tracked action items, and played a key role in maintaining a positive team atmosphere. I contributed to quality control by reviewing and editing individual in-depth and progress reports for 4 members, including the Midterm Report, GANTT chart and Weekly Reports. I actively participated in finding and approving reference papers, ensuring a strong foundation in research. Furthermore, I monitored my own progress and milestones, addressing challenges promptly, and supported my teammates when needed. Overall, my multifaceted contributions significantly contributed to our team's success.
  + In addition to these tasks, I and the deputy leader have organized the weekly meetings regularly. I actively participated in team discussions, bringing valuable insights and expertise to our decision-making processes. I provided guidance and support to my fellow team members, helping them overcome challenges and improve their contributions. My dedication and commitment were pivotal in ensuring our project tasks were completed effectively, and our team operated cohesively.
  + Individual Report Link: [Krupaben\_Kothadia\_Individual\_Report.docx](https://docs.google.com/document/d/1zjFC6-cTxe6-dnSt7yVXMjjZAqcZ400L/edit?usp=drive_link&ouid=107121383631492907164&rtpof=true&sd=true)
* Member Name: [Gautham Vijayaraj](mailto:gvijaya6@asu.edu)
  + I played a pivotal role in various aspects of our project, contributing significantly to its success. Specifically, I took charge of the preparation of our Midterm Report, ensuring that it was both comprehensive and well-structured. Additionally, I consistently evaluated and approved our weekly reports, ensuring they met our quality standards.
  + Furthermore, I had the opportunity to create and maintain the GANTT chart twice. In line with this, I also helped the group leader to assign tasks within our group members and followed up on the tasks for enhanced efficiency.
  + Moreover, I authored a summary report based on four reference papers, synthesizing their key insights. In addition to this, I conducted extensive research on privacy preserving data mining techniques, focusing on privacy and data mining techniques tailored for social media datasets. This research effort added valuable insights to our project's data processing strategy.
  + I actively contributed to our research efforts by actively participating in the process of identifying and approving reference papers. This helped establish a strong foundation for our project within the realm of research and literature.
  + Apart from these, the group leader and I organized the weekly meetings regularly. I actively participated in team discussions, bringing valuable insights and expertise to our decision-making processes.
  + Individual Report Link: [Gautham\_Vijayaraj\_Individual\_Report.doc](https://docs.google.com/document/d/1157cX7Q1fpgZ8voOSeTpuP_VKo93W7Ao/edit?usp=drive_link&ouid=116067103181768245745&rtpof=true&sd=true)
* Member Name: [Avani Mundra](mailto:amudra@asu.edu)
  + I initiated the project by conducting an extensive literature review focused on "Detection of Suspicious Activities in Social Media." This involved thorough web research to identify pertinent papers addressing information assurance and security aspects related to our project.
  + In addition, I actively contributed to crafting the project plan. My involvement included identifying and assessing various risks associated with the project. This process also entailed a meticulous review of the project's scope and objectives to ensure comprehensive coverage of the core ideas.
  + Within my designated research domain, "Data Mining Techniques to Handle Multimodal Data," I conducted in-depth investigations. I curated a selection of relevant research papers to enrich our reference list and enhance our project's knowledge base.
  + Over the subsequent weeks, I diligently prepared weekly progress reports, meticulously documenting the tasks accomplished during each week. These reports included concise summaries of the papers I had reviewed.
  + Furthermore, I produced three comprehensive reports, each focusing on a different research paper. These reports provided a detailed overview of each paper's objectives, the methodology employed, the results achieved (evaluated using various metrics), and concluded with discussions on the paper's findings and potential future research directions.
  + I actively engaged in the evaluation process, providing constructive feedback on [Yeshwanth Reddy Chennur](mailto:ychennur@asu.edu)’s individual and in-depth reports, ensuring the quality and substance of these reports met our project's standards.
  + As part of our project's milestones, I prepared a Midterm Review Report, consolidating all completed tasks to date, showcasing our project's progress.
  + Additionally, I contributed to the Group Midterm Review Report, offering insights into my individual contributions and summarizing the research papers I had reviewed, alongside a record of tasks completed thus far.
  + I took responsibility for preparing the Group Weekly Report, encompassing a comprehensive overview of tasks assigned to each team member for the current week, meeting notes, links to our Gantt Chart, and references to individual progress and in-depth reports.
  + Individual Report Link: [Avani\_Mundra\_Individual\_Progress\_Report.docx](https://docs.google.com/document/d/1keNeP4SwsfaxhxgHCnZAwNZLAwzzWoUT/edit?usp=sharing&ouid=112867995748295868322&rtpof=true&sd=true)
* Member Name: [Anuranjan Dubey](mailto:adubey37@asu.edu)
  + Initiated project by finding relevant machine learning papers on social media suspicion detection.
  + Actively engaged in team meetings and shared research tools. Reviewed provided papers and prepared Weekly Report.
  + Began crafting individual in-depth and progress reports, focusing on machine learning methods.
  + Contributed questions to Group Leader and Deputy Leaders for TA clarification. Provided input for project plan.
  + Continued developing personal reports and assessed team member's work quality.
  + Led the search for relevant datasets and reviewed data mining techniques as well.
  + Contributed to GANTT chart preparation.
  + Identified an irrelevant paper and found a replacement.
  + Involved in midterm report preparation, weekly report and contributing to review, progress, and in-depth reports.
  + Individual Report Link: [Anuranjan\_Dubey\_Individual\_Progress\_Report.docx](https://docs.google.com/document/d/1iDjZoZck6B1N0G3hOK1FSsX9b6ojbBMz/edit?usp=drive_link&ouid=101870164252766530632&rtpof=true&sd=true)
* Member Name: [Sangeeth Santhosh](mailto:ssantho9@asu.edu)
  + Found research papers which are relevant to the main topic of the project - Detection Process of Suspicious Activities on Social Media Using Data Mining and Machine Learning.
  + Collaborated with the group to prepare the project plan. Involved mainly in the project scope section of the project plan.
  + Identified the specific research topic that I would work on for the whole course of the project - Model Evaluation.
  + Took meeting notes of project discussion held via Zoom for the week ending 4th September 2023.
  + Contributed to preparation of a weekly report of the project for the week ending 4th September 2023.
  + For each week, I prepared an individual in-depth report and individual progress report.
  + Evaluated the individual progress report and individual in-depth report of a team member, thus aiding in improving the quality of the project.
  + Prepared the individual midterm report and aided in the completion of the midterm report.
  + Individual Report Link: [Sangeeth Santhosh\_Individual\_Progress\_Report.docx](https://docs.google.com/document/d/1BmYyfuvH53TX1Y6lTpXB3_9a1KqHcW0f/edit?usp=drive_link&ouid=100662470547641371728&rtpof=true&sd=true)
* Member Name: [Rahul Nayak](mailto:rrnayak@asu.edu)
  + In the first week I was assigned the task of doing a literature survey for the topic “Detection of Suspicious Activities on Social Media Using Data Mining and Machine Learning” .I selected the research topic types of suspicious/Malicious activities that happen in social media, Hence came up with 4 research papers for the same.I was given the opportunity to create and maintain the GANTT chart in this week, which aided in the effective management of project timelines. I was also given the task of evaluating Avani Mundra’s Weekly progress report and In depth report. This was a recurring task that I had to complete every week.
  + In the second week,I prepared my progress report and in depth report for the paper Suspicious Behavior Detection Current Trends and Future Directions.
  + In the third week I prepared my progress report and in depth report for the paper Malicious Profile Detection on Social Media: A Survey Paper. I plan on doing a review on different types of malicious/suspicious activity every week. This week, I reviewed malicious profiles on Social media.
  + In the fourth week I was responsible for developing the weekly report. I also prepared a progress report and in depth report for the paper Machine Learning Based Twitter Spam Account Detection A Review. This week the type of malicious activity was spam accounts on twitter and detection using machine learning techniques.
  + In the fifth week,I prepared my progress report for the paper Detecting spammers on social networks and also was responsible for preparing the midterm review report.
  + Individual Report Link: [Rahul\_Nayak\_Individual\_Report.docx](https://docs.google.com/document/d/1Ddglynobb-b7-np-6Fs6hHTGmQhdk75P/edit#heading=h.ex4ljgcoifmx)
* Member Name: [Yeshwanth Reddy Chennur](mailto:ychennur@asu.edu)
  + Initially, I gathered many research papers related to the project and went through them, then finalized 5 papers which are most relevant to the project title.
  + Then, I actively participated in preparing the project plan and the weekly report. I was incharge of doing the literature review, project scope. Also, collectively worked on the entire project plan preparation.
  + Had zoom meetings every week and discussed the week plan and distributed the work equally among us.
  + From the next week, I started studying the research papers. In the first week, I started reading the paper, “[Towards automated real-time detection of misinformation on Twitter](https://drive.google.com/file/d/12Lm3n-1HkhbutIC-yg4k6kkTWQTXhiwn/view?usp=drive_link)” and made an individual progress report and in-depth report on the paper with the key points and the outcomes from the paper.
  + In the second week, I started reading the paper, “[A Trusted Social Network Using Hypothetical Mathematical Model and Decision- Based Scheme](https://drive.google.com/file/d/15-gauHeasPzFDcXhp6a1OyeXj3vuIdrT/view?usp=drive_link)” and finished it the next week. Also, I made an in-depth report of the paper in the third week.
  + Responsible for evaluating the progress reports and in depth reports of Justin Young every week.
  + Contributed towards the Midterm Review Report preparation and completed the Individual contribution in the midterm review report.
  + Individual Report Link: [Yeshwanth\_Reddy\_Chennur\_Individual\_Report](https://docs.google.com/document/d/1Z7il4cOV3Ji0XNmYzUvsy0hDiTul8WP4/edit?usp=drive_link&ouid=100701464624984336471&rtpof=true&sd=true)
* Member Name: [Justin Young](mailto:jtyoun15@asu.edu)
  + At the start of the project, I gathered research papers to provide a general overview of our project topic “Detecting Suspicious Activity on Social Media using Machine Learning and Data Mining”.
  + In the second week I was assigned the focus topic of “Ensuring model security”, and gathered 4 research papers related to that topic. I also helped prepare the Gantt chart for that particular week. I also prepared an in-depth report on the research paper “Behavior-Based Malware Detection System Approach For Mobile Security Using Machine Learning.
  + In the third week I prepared an in-depth report for the research paper “Comparative Analysis of Machine Learning Models in Computer Network Intrusion Detection” and helped prepare the weekly report for that week.
  + In the fourth week I prepared an in-depth report for the research paper “Cooperative Machine Learning Techniques for Cloud Intrusion Detection”.
  + In the fifth week I prepared an in-depth report for the research paper “Machine Learning-based web security intrusion detection system” and helped prepare the weekly report for that week.
  + Overall my studies for this project so far focuses on using machine learning models to ensure security on various different platforms such as mobile devices and cloud systems, and comparing the performance of various models to determine which is the most effective and any downfalls they may have.
  + Individual Report Link: [Justin\_Young\_Individual\_Report](https://docs.google.com/document/d/121ED9I0MXZwvfzIB1EJS0m4bWfJFF2hQ/edit?usp=drive_link&ouid=117665874805656519311&rtpof=true&sd=true)

Next Steps

* Every team member will come up with an individual in-depth and progress report every week on their remaining papers.
* Every team member will go through all the Not so Important Papers to have more knowledge about the work done by others in the domain of our project.
* We will come up with the Final report.

## Submission Details

**Name of the member that prepared the report:** [Krupaben Kothadia](mailto:kkothadi@asu.edu)**,** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu)**,** [Avani Mundra](mailto:amudra@asu.edu)**,** [Justin Young](mailto:jtyoun15@asu.edu)**,** [Anuranjan Dubey](mailto:adubey37@asu.edu)**,** [Rahul Nayak](mailto:rrnayak@asu.edu)**,** [Sangeeth Santhosh](mailto:ssantho9@asu.edu)**,** [Yeshwanth Reddy Chennur](mailto:ychennur@asu.edu)

**Name of the member that submitted this report:** [Krupaben Kothadia](mailto:kkothadi@asu.edu)  
  
  
**The report is approved by:** [Krupaben Kothadia](mailto:kkothadi@asu.edu)[Gautham Vijayaraj](mailto:gvijaya6@asu.edu)