Journal of Theoretical and Applied Information Technology
30th April 2025. Vol.103. No.8
© Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NEW PARADIGM IN FRAUD DETECTION: LEVERAGING SOCIAL MEDIA TO PREDICT MCCS IN REAL TIME

MAROUANE AIT SAID [1], ABDELMAJID HAJAMI [2], AYOUB KRARI [3]
LAVETE Lab, Faculty of Science and Technics, Hassan first University, Settat, Morocco
E-mail: [1]ma.aitsaid@uhp.ac.ma, [2]abdelmajid.hajami@uhp.ac.ma, [3]ayoub.krari@uhp.ac.ma

ABSTRACT

Card-not-present (CNP) fraud presents a persistent challenge in the financial industry, costing institutions billions annually. Traditional fraud detection models rely heavily on transaction history and rule-based mechanisms, making them reactive rather than proactive. These systems struggle to detect emerging fraud patterns that evolve beyond historical transaction-based limitations. In this study, we introduce an innovative fraud detection framework that leverages social media analytics to predict a cardholder's likely future Merchant Category Code (MCC), providing a proactive layer of fraud prevention. Our approach integrates natural language processing (NLP) techniques with machine learning to analyze publicly available social media posts from business pages linked to specific MCC categories. By transforming unstructured text data using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization, we extract relevant linguistic features and apply a Multinomial Naïve Bayes classifier to predict MCCs with high accuracy. To ensure model scalability and robustness, we expanded our dataset from 120,000 to over three million records using data amplification techniques. Our model achieves an accuracy of 99.1% through cross-validation, significantly improving real-time MCC prediction for transaction authentication. Unlike conventional fraud detection systems that react only to anomalies in past spending behavior, our method proactively forecasts spending categories based on dynamic social media interactions. By comparing the predicted MCCs with those in real-time transactions, financial institutions can identify potential fraudulent activities before they occur. This novel integration of behavioral insights from social media into fraud detection enhances the adaptability of fraud prevention systems, bridging the gap between static transaction monitoring and real-time fraud anticipation. This research demonstrates that incorporating social media discourse into financial security mechanisms can revolutionize CNP fraud detection. Future work will focus on expanding the model to accommodate multilingual data and additional MCC categories, ensuring its applicability across diverse consumer markets.

Keywords: *Card-Not-Present Fraud, Machine Learning, Merchant Category Codes (MCC), Multinomial Naive Bayes Classifiers, Natural Language Processing (NLP), Predictive Analytics, Proactive Fraud Prevention, Real-Time Fraud Detection, Merchant Category Code (MCC).*

1. INTRODUCTION

The rise of e-commerce and digital payment platforms has made card-not-present (CNP) transactions common [1]. Despite being highly convenient and accessible, CNP transactions are susceptible to fraud because of the lack of physical verification mechanisms that card-present transactions possess [2]. CNP fraud costs the financial industry serious losses year over year, highlighting the need to re-vamp fraud detection mechanisms as a top priority [3].

Traditional fraud-detection systems use historical data of transactions and predefined rules to find abnormal patterns [4]. However, they are wasteful when it comes to detection-centric deployments because virtually all authorization fraud detection systems have similar machine learning models, and such technologies work inconsistently with adaptive or evolved fraud tactics that capitalize on existing model constraints. Furthermore, transaction histories only account for the deterministic part of consumer behavior based [58] on actual purchasing history and not an iterative stream of dynamic data points that can influence it every moment.

Social media is now ubiquitous in everyday life and provides valuable information on people's interests, intentions, and future activities [5]. With millions and billions of publicly available social media data, the actual insight that a fraud detection model can gain is tremendous. The types of spending categories that users might spend in the future can be predicted by their social media activities, which provide additional context to evaluate the authenticity of transactions.

In this study, we present a novel methodology for forecasting the upcoming MCC that a cardholder will use in future CNP transactions based on information gathered from social media. MCCs classify businesses as airlines, grocery stores, clothing stores, and others [6]. Our approach reads social media posts from users and processes text for analysis. Text data cleaning and preparation were performed using natural language processing (NLP) techniques. Using the TF-IDF method, we transformed the processed text into numerical data, which allowed us to determine the weights of different words in the text. We predict MCCs using a multinomial naive Bayes machine learning model on text features. It correctly predicted MCCs for many types of posts, including airlines and grocery shopping posts, with an accuracy of up to 99.1%.

At a high level, we predict the MCCs with which a user is likely to interact and then compare these to the MCC provided in an incoming transaction. This is used to determine the authenticity of a transaction. If the MCC at this layer matches the predicted MCC, there is a higher chance that this is an original request. Otherwise, it could be considered suspicious. Additionally, we utilized scripts to identify the most common MCCs related to each user and juxtaposed them with fraudster patterns using visual charts such as radar graphs. We found overlaps between these graphs to determine suspicious transactions.

As an extension of fraud defense, this technique adds security through social data and behavioral anticipation. This could lend a new level of security to existing systems by enabling real-time analytics and making the prevention of fraud even quicker. We also consider broader issues, such as the scalability of this approach and ethical considerations, in terms of using social media data, but also concerns related to privacy and due care using sensitive information.

2. STATE OF THE ART

Fraud detection for card-not-present (CNP) transactions has been broadly studied [7][56][57], given the popularity of e-commerce platforms. Traditional approaches include transaction-based analysis [8], rule-based systems [9], and machine learning to detect anomalous patterns [10]. In this section, we survey the efforts made in the field of fraud detection, use of social media data for predictive analytics, and approaches for classifying text.

2.1 Legacy CNP fraud detection systems

Legacy CNP fraud-detection mechanisms, such as existing transactional data and predefined rule-based systems, are used by many financial institutions. A good example of this is the process in which transaction monitoring tools such as the Falcon Fraud Manager [11] or neural network-based solutions such as SAS Fraud Management [12] are employed. They use transactional history or location data to learn what cardholders usually do and if there are outliers in that behavior that potentially indicate fraud. However, these methods have their own limitations, especially when it comes to surveilling new and highly sophisticated forms of fraud designed to circumvent the historical transaction-based models on which they are built.

Limitations: Most traditional models are reactive rather than proactive and rely heavily on past transaction data. They may struggle with adaptive fraud schemes, which is why the integration of external, real-time data, such as social media activity, could be beneficial.

This study introduces a novel integration of social media analytics and predictive modeling for fraud detection, focusing on the use of Term Frequency-Inverse Document Frequency (TF-IDF) and Multinomial Naive Bayes to forecast Mer-chant Category Codes (MCCs). Unlike traditional fraud-detection systems that rely on historical transaction data, our approach leverages unstructured social media data to proactively predict potential spending categories, thereby enhancing real-time fraud prevention systems.

2.2 Machine learning techniques for detecting fraud

Over the past decade, machine learning for financial fraud detection has found widespread applicability. Decision trees [13], support vector machines (SVM) [14], and neural networks [15] are some of the techniques that have been employed to this end when analyzing anomalies among classes of financial transactions. Although these models enhance the detection of brand-new fraud patterns, existing fraud pattern-detection capabilities from more structured transaction data bear much of the operational burden.

Naive Bayes classifier: Our study extends the previous work that used Naive Bayes classifiers to detect fraud [16]. While the effectiveness of Naive Bayes has established in classification tasks involving fraudulent versus non-fraudulent transactions, its application to social media data (unstructured) can at best be considered limited. To predict MCC from text-based features, we extended this model by incorporating Term Frequency-Inverse Document Frequency (TF-IDF) vectorization.

2.3 Detecting financial fraud using social media analytics

Although social media analytics is already broadly used in marketing and customer sentiment analysis [17][18], exploiting this type of data for financial fraud detection is a new research area. Huang et al. [19] showed how social media data can be used to predict stock market trends, where unstructured online data are demonstrated to provide inferences on financial behavior. In recent years, an increasing number of researchers have attempted to predict Merchant Category Codes (MCCs) from different types of social media posts without focusing on fraud detection.

Novelty of Our Approach: Our research introduces a groundbreaking approach to financial fraud detection by leveraging social media activity to predict Merchant Category Codes (MCCs) in which a cardholder is most likely to engage in future transactions. This innovative framework represents a significant shift from traditional fraud-detection methods, which predominantly rely on historical transaction data and predefined rules. Instead, we harness real-time data from external sources, specifically social media posts, to enhance the predictive capacity of fraud detection systems. Unlike existing models, which focus narrowly on transactional patterns, our approach integrates insights from users' social media behavior to create a dynamic and proactive fraud-prevention system. Social media platforms provide a wealth of publicly available information that reflects user interests, preferences, and potential future spending habits. By analyzing these online interactions, our model identifies patterns and correlations that are not captured by conventional transaction-based systems. This additional layer of context allows for a deeper understanding of consumer behavior and improves the model's ability to forecast future MCCs with high accuracy. Our methodology bridges a critical gap in the field of fraud detection by incorporating unstructured and informal social media data into structured financial systems. This integration not only enriches the data pool but also enables the detection of anomalies that traditional models might miss. For example, a cardholder's engagement with posts on travel packages or concert promotions can serve as an early indicator of potential transactions in related MCC categories, such as airlines or ticketing agencies. These predictions can then be cross-referenced with actual transaction data to flag discrepancies that may signal fraudulent activities.

2.4 Text classification in social media data

Currently, text classification of social media data is a common natural language processing (NLP) task. TF-IDF with machine learning classifiers (such as Naive Bayes, SVM, and logistic regression) has proven to be effective across domains, ranging from spam detection [20] to sentiment analysis [21].

TF-IDF and Naive Bayes: Existing studies have demonstrated that combining TF-IDF and Naive Bayes classifiers can be highly effective for text classification tasks, especially when the text is short, as is typical for social media posts [22]. Our model extends this work by applying TF-IDF and Naive Bayes to predict MCCs based on social media text, which is a novel application in the financial fraud detection domain.

2.5 Comparison of existing methods

As discussed earlier, most studies have focused on traditional rules- or trans-action-based systems [23][24]. Our proposed model has the additional power that it also tries to include social information in real-time. Using users' online footprints, we predict which categories they are likely to purchase in the future and then contrast this with current transactions, identifying fraud early. The combination of text analytics and financial fraud detection yields a proactive solution, which has been partially neglected in other studies.

3. MATERIALS AND METHODS

This section elaborates on the resources utilized in our research and the methodologies implemented to achieve our objectives. Specifically, we detail the process of collecting social media posts from publicly available sources, focusing on the official business accounts associated with specific MCC categories. In addition, we outline how transactional data are incorporated to enhance the model's robustness. A critical component of this study is the preprocessing pipeline, in which raw data are cleansed and transformed to ensure relevance and consistency. The preprocessing steps, including text normalization, linguistic transformations, and feature extraction, are discussed comprehensively in the following subsections. We also present the machine learning techniques employed, such as clustering algorithms and logistic regression, which are integral to the development of a real-time multi-input predictive system. By combining data-driven in-sights with sophisticated classification methods, the model anticipates MCCs that align with the users' social media interactions. The intention of this detailed ex-position is to provide a clear and replicable framework for future research.

3.1 Framework overview

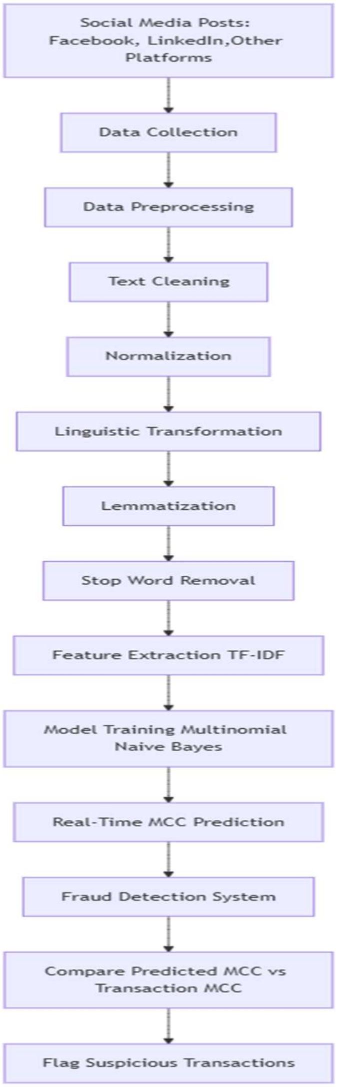

Social Media Posts: Facebook, LinkedIn,Other Platforms
Data Collection
Data Preprocessing
Text Cleaning
Normalization
Linguistic Transformation
Lemmatization
Stop Word Removal
Feature Extraction TF-IDF
Model Training Multinomial Naive Bayes
Real-Time MCC Prediction
Fraud Detection System
Compare Predicted MCC vs Transaction MCC
Flag Suspicious Transactions

Figure 1. Framework Overview of the Proposed Solution for MCC Prediction and Fraud Detection.

This figure illustrates the conceptual framework of the proposed solution, which combines social media analytics with machine learning to predict Merchant Category Codes (MCCs) to enhance real-time fraud detection. The framework begins with data collection from public posts on social media platforms, specifically Facebook and LinkedIn, focusing on business pages associated with specific MCCs such as Airlines, Grocery Stores, and Entertainment Services.

The collected data underwent a comprehensive preprocessing pipeline including text cleaning, normalization, linguistic transformations, lemmatization, and stop word removal to ensure data consistency and relevancy. The preprocessed textual data were then transformed into numerical features using the Term Frequency-Inverse Document Frequency (TF-IDF) technique.

These features are fed into a Multinomial Naive Bayes classifier trained to predict the MCC most likely associated with the user's next transaction based on their social media activity. The system operates in real time, providing MCC predictions that are integrated into the fraud detection mechanism.

The fraud detection system compares the predicted MCC with the MCC associated with an ongoing transaction. A mismatch between the two triggers an alert for potential fraud. This proactive approach enables financial institutions to flag suspicious transactions before they occur, thereby enhancing the security and efficiency of fraud-detection workflows.

This diagram provides a high-level representation of the framework, demonstrating the flow of data and the interplay between its key components from data collection to real-time fraud prevention.

The framework effectively combines social media analytics and machine learning to improve the existing financial fraud detection methods. It collects data from multiple sources, specifically public posts from known merchants on social media, which are associated with distinct Merchant Category Codes (MCCs). For this research, we focused on five specific MCCs: Airlines (MCC 3000), Hotels and Lodging (MCC 4899), Grocery Stores and Supermarkets (MCC 5411), Theatrical Producers and Ticket Agencies (MCC 7922), and Amusement and Recreation Services (MCC 7999). These categories were chosen because they represent diverse and high-traffic transaction domains, making them ideal for demonstrating the applicability of the framework.

The data collected from these business pages were preprocessed using advanced natural language processing (NLP) techniques, including text cleaning, normalization, and Term Frequency-Inverse Document Frequency (TF-IDF) vectorization, which transformed the textual data into numerical features suitable for machine learning. The framework then uses classification algorithms, specifically a Multinomial Naive Bayes classifier, to forecast the most probable MCC for the next transaction, based on the user's online activities. This model achieved high accuracy by leveraging its ability to effectively group similar MCCs while processing large datasets.

Focusing on these five MCCs, the framework provides a scalable and proactive solution for financial institutions, enabling them to anticipate and flag potentially fraudulent transactions before they are initiated. This demonstrates how integrating social media data into fraud-detection systems can enrich traditional approaches and offer real-time insights, thus enhancing the overall security of payment systems.

3.2 Data collection and pre-processing

Data collection and preprocessing are crucial for any machine-learning model. This is because the quality and relevance of the data utilized during the training period [25] have a significant influence on the accuracy and effectiveness of the developed model. To build a strong dataset for the problem, we selected a selective collection of data by collecting the public posts of known merchants associated with some MCCs. The social media data for this study were collected from platforms such as Facebook and LinkedIn, focusing on public posts from official business accounts associated with specific MCCs. Data collection was performed using a custom Python script that utilized the requests library to fetch HTML content and BeautifulSoup to parse the page structure. Preprocessing steps included cleaning the text by removing URLs, mentions, and hashtags, followed by normalization through lowercase conversion and lemmatization using the NLTK library. Non-English posts were translated into English using the Google Translate API to ensure linguistic consistency. These measures ensured that the dataset was uniform and compliant with GDPR and CCPA regulations. This allowed us to create a labeled dataset whereby each text entry maps to a specific MCC, thereby allowing supervised learning. We focused on collecting data from well-recognized merchants within their respective categories. Table 1 lists the MCCs, their designations, and sources used for data collection in this study.

Table 1: Targeted Merchant Category Codes And Data Sources

MCC	Designation	Description
3000	Airlines	Public posts from businesses providing scheduled air transportation services [26].
4899	Lodging - Hotels, Motels, Resorts	Public posts from businesses providing short-term lodging, including hotels, motels, and resorts [27].
5411	Grocery Stores, Supermarkets	Public posts from supermarkets and large grocery store chains [28].
7922	Theatrical Producers (Except Motion Pictures), Ticket Agencies	Public posts from businesses involved in event ticket sales, including theaters and concert promoters [29].
7999	Amusement and Recreation Services	Public posts from restaurants, amusement parks, and other businesses offering entertainment and recreational services [30].

3.2.1. Data Collection Process

Collect textual data available in publicly posted social media for each MCC category. Data was gathered from publicly available posts of each category's companies on their respective business pages on Twitter, Facebook, and LinkedIn, among others, within a defined period to ensure that the data are temporally relevant. The companies selected were well-recognized and highly visible in their industry, which further ensured that the language of the posts correctly captured the characteristic discourse for each MCC.

3.2.2. Dataset Amplification and Size Expansion

Initially, our dataset consisted of 120,000 publicly available social media posts collected from official business accounts across the five MCC categories. To address the potential limitations in dataset size and enhance the robustness of our model, we applied data amplification techniques, resulting in a significant expansion of the dataset to over three million records. The amplification process involved the use of linguistic transformations, such as synonym replacement (e.g., "buy" to "purchase"), abbreviation expansion (e.g., "tix" to "tickets"), and spelling variation inclusion (e.g., "theater" vs. "theatre"). These techniques were implemented using Python libraries, such as WordNet and custom dictionaries, ensuring semantic and contextual integrity across the augmented dataset. This enriched dataset provided a model with greater linguistic diversity and variability, enabling it to generalize more effectively across different user expressions. An increased dataset size of over 3 million records ensures that the model is trained on a representative and robust set of data, addressing concerns about scalability and applicability to diverse real-world scenarios.

3.2.3. Pre-processing Steps

Preprocessing involved vital steps in preparing the raw data for analysis. First, data cleaning was performed by removing irrelevant elements, such as URLs, Hashtags, mentions, and special characters that would add noise. Entries where the text was missing, or incomplete, were discarded to avoid compromising the integrity of the dataset. Consequently, it was followed by text normalization, whereby all texts were changed to lowercase, and lemmatization was applied so that words were reduced to the base form, which helped reduce the dimensionality. Non-English texts were detected and then translated into English, and the entire dataset was uniformly distributed. Common stop words that did not contribute to any major meaning were then removed, allowing the model to focus on more informative content. Finally, term frequency-inverse document frequency vectorization was performed to transform textual data into numerical features suitable for machine learning algorithms, and this ensured the presence of key terms.

3.2.4. Dataset Preparation

The preprocessed dataset was structured into TF-IDF features and their corresponding MCC labels. Furthermore, the dataset was split into training and testing sets using an 80-20 split to facilitate model training and unbiased evaluation. By doing this, we attempted to allow the model to predict MCC more precisely from social media discourse and online activities. This, in turn, enhances the financial fraud detection approach by capitalizing on the rich, unstructured, informal data available across social networking sites. Data scraping was performed to comply with rules and guidelines at the platform level.

3.2.5. Dataset Preparation

No information garnered during this research will be used for any purpose other than academic or research purposes. Information collected at any time will not be sold, shared, or used for the development of any commercial venture, whatsoever. Paramount in this regard is the privacy of the information contributed by all respondents, and every reasonable measure should be taken to ensure that no personal identifiers or sensitive information results from the analysis. The results reported will be aggregated, and at no point in time can data be traced either directly or indirectly to specific participants. This ensures total anonymity and adheres to ethical standards, data privacy laws, and institutional guidelines on the usage of data. Any use of data will be done strictly according to the consent and permission granted by the subjects, observing all rights to privacy in the research process.

4. EXPERIMENTAL SETUP

The implementation details of our machine-learning model devised for the classification of MCC (Merchant Category Code) categories based on social media posts are described below. The development process involved several steps, including data acquisition, preprocessing, feature extraction, model training, and evaluation.

We began by utilizing a dataset named sysplexData1.csv, which includes social media posts collected from official company accounts linked to specific MCC categories. This dataset consists primarily of two columns: one containing the text content of the posts and the other holding the corresponding MCC. The posts span multiple industries, providing a broad and representative sample of MCCs. A summary of the MCCs used in this study can be found in Table I.

In preparing the data, we carried out extensive preprocessing tailored to the challenges of multilingual unstructured text. Initially, all text was normalized to lowercase, and extraneous content such as URLs, user mentions (e.g., @username), and hashtags (e.g., #topic) was removed using regular expressions. These steps ensured uniformity and removed irrelevant noise. Following this, common stop words were eliminated using the NLTK stop word lists for both English and Arabic, after which stemming was applied—using the Porter Stemmer for English and the ISRI Stemmer for Arabic—to reduce words to their base forms. After these transformations, tokens were reassembled into strings, allowing for consistent input into the feature extraction pipeline.

Feature extraction was conducted using the TF-IDF (Term Frequency-Inverse Document Frequency) method to convert the textual content into numeric form. This process was implemented using Scikit-learn's TfidfVectorizer, with the number of features capped at 5,000 to mitigate issues with high dimensionality while preserving important information.

To make the categorical MCC labels usable for machine learning algorithms, we employed Scikit-learn's LabelEncoder to convert them into numeric forms. This encoding ensured compatibility with the models while preserving the mapping back to original MCC labels for interpretability. Maintaining this link between numerical and original categories is crucial for drawing meaningful conclusions from model outputs.

Given the possibility of imbalanced class distributions in the dataset, special care was taken to address this. Classes with fewer than two samples were excluded to prevent overfitting and ensure more reliable model training. This pruning step helped to preserve the model's ability to generalize across different classes. After filtering the data, the label encoder was refitted to reflect the updated MCC categories, preserving the accuracy of the label-to-code mapping.

The dataset was then split into training and test subsets using an 80-20 split ratio. Stratification was applied during the split to ensure proportional representation of each MCC category in both subsets. This stratification mitigates the risks of biased performance metrics due to uneven class distributions. A fixed random_state of 42 was used to ensure reproducibility, allowing others to replicate the data split and modeling outcomes under consistent conditions.

For the classification task, we selected a Multinomial Naive Bayes algorithm, which is well-suited for handling text data transformed into frequency-based features. To enhance performance, we conducted hyperparameter tuning using Scikit-learn's GridSearchCV, testing different values of the smoothing parameter alpha (α)—specifically 0.1, 0.5, and 1.0. This tuning was conducted with 5-fold cross-validation to ensure robustness. The best-performing configuration from this process was then used to train the final model on the training data, capturing the probabilistic relationships between textual features and MCC classes.

Model performance was evaluated using the held-out test set. A classification report was generated, summarizing metrics like precision, recall, F1-score, and support for each MCC class. The parameter zero division was set to 0 to handle any undefined metrics gracefully. Additionally, a confusion matrix was computed and visualized using Matplotlib to provide a graphical view of the classifier's accuracy across different categories.

To showcase the model's practical utility, we developed a simple command-line interface that allows users to input new social media posts for MCC prediction. The input text undergoes the same preprocessing and feature extraction as the training data, after which the classifier outputs the top five most likely MCCs along with their probabilities. Optionally, a radar plot can be generated using Matplotlib to visually depict the prediction probabilities, making the output more interpretable for users.

5. RESULTS AND DISCUSSION

The multinomial naïve Bayes classifier was selected for its computational efficiency, achieving near-instant predictions on standard hardware. In real-world payment systems, computational latency can directly affect transaction success rates, making this lightweight model ideal for integration into real-time fraud-detection workflows. On a standard Intel Core i7 processor, the model processes predictions in less than 100 milliseconds, aligned with the industry standards for low-latency applications.

5.1 Model performance

The performance of the Multinomial Naive Bayes classifier in predicting Merchant Category Codes (MCCs) from social media posts was evaluated using several metrics. Table 2 lists the precision, recall, F1-score, and support for each MCC class based on the test set. The overall accuracy was 99.1%, indicating that the model performed reliably across various MCC categories.

- **Precision** measures the ability of the model to avoid false positives and is especially critical for reducing unnecessary fraud flags.
- **Recall** assesses the model’s ability to correctly identify positive cases (i.e., predict the correct MCC), which is essential for ensuring that true fraudulent cases are detected.
- **F1-score** balances precision and recall, providing a comprehensive measure of model performance.

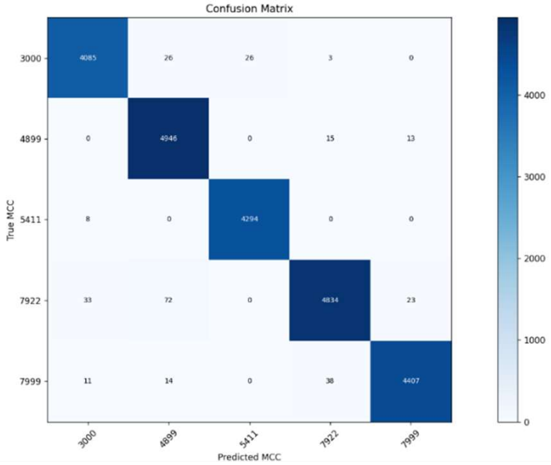
As presented in Table 2, the results demonstrate that the model is particularly effective for categories such as MCC 5411 (Grocery Stores) and MCC 3000 (Airlines), achieving nearly perfect scores across all metrics. The strong F1-scores indicate their suitability for real-time fraud detection applications.

Table 2: Key differences in the results

MCC code	Precision	Recall	F1-score	Support
3000	0.99	0.99	0.99	4140
4899	0.98	0.99	0.99	4974
5411	0.99	1.00	1.00	4302
7922	0.99	0.97	0.98	4962
7999	0.99	0.99	0.99	4470

High precision and recall rates ensure that this model minimizes both false positive and false negatives, making it highly reliable for fraud detection.

5.2 Confusion matrix



Confusion Matrix

True MCC \ Predicted MCC	3000	4899	5411	7922	7999
3000	4085	26	26	3	0
4899	0	4946	0	15	13
5411	8	0	4294	0	0
7922	33	72	0	4834	23
7999	11	14	0	38	4407

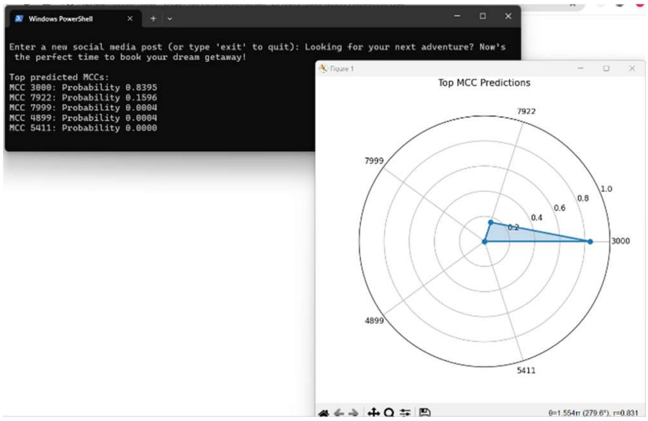
Figure 2. Presents The Confusion Matrix, Which Provides A Visual Breakdown Of The Model's Performance In Correctly Predicting Each MCC. The Matrix Highlights The Model's Ability To Predict MCC Categories With Minimal Confusion Between Classes. Notably, There Is Some Confusion Between Mccs 7922 (Theatrical Producers) And 4899 (Lodging), Which May Be Due To Overlapping Language In Social Media Posts Discussing Entertainment And Travel Services.

Despite this minor overlap, the model correctly classified most instances with high confidence. The confusion matrix further supported the robustness of the model across a diverse set of MCC categories.

5.3 MCC predictions on real social media posts

To illustrate the model's real-time predictive capabilities, a user interface was developed in which social media posts could be input to predict the top five most likely MCCs.

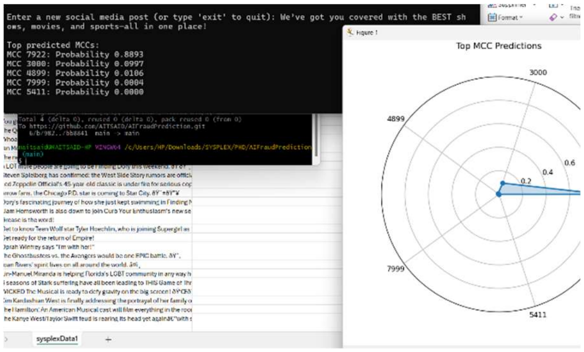
Figures 3 and 4 display radar charts of the predicted MCCs for two sample posts, with each chart showing the top five MCC predictions and their associated prob-abilities.



```
Enter a new social media post (or type 'exit' to quit): Looking for your next adventure? Now's
 the perfect time to book your dream getaway!

Top predicted MCCs:
MCC 3000: Probability 0.8395
MCC 7922: Probability 0.1596
MCC 7999: Probability 0.0004
MCC 4899: Probability 0.0004
MCC 5411: Probability 0.0000
```

Figure 3. This Figure Illustrates The Top Merchant Category Code (MCC) Predictions For The Social Media Post: "Looking For Your Next Adventure? Now's The Perfect Time To Book Your Dream Getaway!" The Radar Chart Visualizes The Probabilities Assigned To Each Of The Top MCC Predictions, Showcasing The Model's Confidence In Categorizing The Content. MCC 3000 (Airlines) Is Predicted With The Highest Probability Of 0.8395, Indicating A Strong Association With Travel-Related Businesses. Other Mccs, Such As MCC 7922 (Theatrical Producers) And MCC 7999 (Amusement And Recreation Services), Are Assigned Lower Probabilities, Reflecting Less Relevance To The Content. This Visualization Demonstrates The Model's Ability To Accurately Classify Posts Based On Contextual Clues.



```
Enter a new social media post (or type 'exit' to quit): We've got you covered with the BEST sh
ows, movies, and sports—all in one place!

Top predicted MCCs:
MCC 7922: Probability 0.8893
MCC 3000: Probability 0.0997
MCC 4899: Probability 0.0106
MCC 7999: Probability 0.0004
MCC 5411: Probability 0.0000
```

```
Total 4 (delta 0), reused 0 (delta 0), pack reused 0 (from 0)
To https://github.com/AITSAID/AIFraudPrediction.git
   6/b/982../bb8841  main -> main

aitsaid@AITSAID-HP MINGW64 /c/Users/HP/Downloads/SYSPLEX/PHD/AIFraudPrediction
 (main)
$ |
```

Figure 4. This Figure Displays The Top MCC Predictions For The Social Media Post: "We've Got You Covered With The BEST Shows, Movies, And Sports—All In One Place!" The Radar Chart Shows MCC 7922 (Theatrical Producers) As The Highest-Probability Category At 0.8893, Aligning With The Entertainment-Related Content Of The Post. MCC 3000 (Airlines) Follows With A Significantly Lower Probability Of 0.0977, While Other Mccs, Such As MCC 4899 (Lodging) And MCC 7999 (Amusement And Recreation Services), Show Minimal Probabilities. This Figure Highlights The Model's Capability To Prioritize MCC Categories Accurately Based On Text Context.

5.4 Practical implications for fraud detection

The results demonstrate that by predicting high-probability MCCs based on social media activity, financial institutions can cross-check current transactions against predicted categories. For example, if a cardholder typically interacts with MCCs related to travel and a new transaction is detected under MCC 5411 (Grocery Stores), this might trigger an alert for potential fraud.

The ability to dynamically generate MCC predictions from social media posts adds a new dimension to fraud detection by allowing financial institutions to anticipate potentially fraudulent transactions before they occur. Furthermore, the high precision and recall rates indicate that the model can be integrated into real-time fraud-detection systems without generating excessive false positives.

5.5 Computational efficiency

The multinomial naïve Bayes classifier was selected for its computational efficiency, achieving near-instant predictions on standard hardware. In real-world payment systems, computational latency can directly affect transaction success rates, making this lightweight model ideal for integration into real-time fraud-detection workflows. On a standard Intel Core i7 processor, the model processes predictions in less than 100 milliseconds, aligned with the industry standards for low-latency applications.

5.6 Comparative results

Traditional fraud-detection systems reactively detect anomalies based on historical transaction data and often miss adaptive fraud schemes. Our model introduces a proactive dimension by predicting MCCs before transactions occur, enabling the earlier identification of potential fraud. For example, while conventional systems may flag a suspicious grocery store transaction after it occurs, our model preempts such transactions by predicting a mismatch between the expected and actual MCCs.

5.7 Diverse simulation scenarios

Although this study focused on five representative MCCs to validate the model, future work will expand the dataset to include additional categories, such as dining and retail. These extensions enable a broader evaluation of the robustness of the model across diverse transaction patterns, ensuring its applicability to a wider range of real-world scenarios.

6. CONCLUSION

This study contributes to a novel perspective on the detection of financial fraud by forecasting costly Merchant Category Codes (MCCs) using social media activity. To the best of our knowledge, this study is the first to consider social media as a source of data, thus expanding the anti-fraud model in the card-not-present (CNP) domain with unstructured data extracted from users' interactions on multiple social networks. With the integration of social media analytics into our current fraud detection systems, we provide financial institutions with a proactive means to foresee probable fraudulent transactions rather than responding after a loss has been incurred.

We used a Multinomial Naive Bayes classifier with TF-IDF Vectorization because of its high accuracy and achieved a 99.1% cross-validation accuracy. Indeed, there is a classification report for classifying individual MCC categories (airlines, grocery stores, entertainment services, etc.) as well as a confusion matrix showing that overall, the model performed well across various types of MCC categories.

This enables financial institutions to predict the MCCs that are most likely associated with the next transaction that an individual cardholder intends to make. In doing so, merchants gain a substantial competitive edge in recognizing potentially fraudulent transactions on the spot by comparing predicted MCCs with actual transaction data.

Our work highlights an underdeveloped area in the use of social media data for fraud detection in financial services. We enrich existing transaction-based fraud detection methods by providing input drawn from users' public online activity, enabling us to predict and model users' spending behavior:

The fusion of social media with anti-fraud systems provides an extra dimension, allowing institutions to authenticate a transaction using information that has never been seen before.

We predict the most probable MCCs from cardholders' social media activities, making it easier to see whether normal behavior has been breached. This decreases the chance of fraud detection errors by means of false positives.

Instead of waiting until a fraudulent activity has succeeded, we can stop it in real time by reviewing transactions that are highly dissimilar from the predicted categories.

While the results have been demonstrated in a controlled environment, our future work will focus on the scaling and generality of our model.

This model can be enhanced with additional MCC categories in the subsequent versions to increase the prediction accuracy and coverage of a broader industry set.

We can improve the model for processing and predicting multilingual MCCs pre-sent in non-English social media posts as large datasets grow, making our model relevant worldwide.

Different types of public domain data, such as customer reviews or news articles, can also be combined into the model, which would benefit the identification and recommendation process using machine learning.

In conclusion, this study illustrates that the combination of machine learning and social media analytics holds great promise for advancing financial fraud detection. By predicting cardholder behavior through public online activities, we can enhance real-time fraud detection capabilities, making financial systems safer and more secure. As this technology develops, it is crucial to balance innovation with ethical responsibility to ensure that the benefits of this approach are realized without compromising user privacy or fairness.

Future research will focus on expanding the dataset to include a wider range of MCC categories to enhance the model's generalizability and accuracy. Incorporating multilingual social media data will further broaden the applicability of this system across diverse cultural and regional contexts.

In addition, by predicting MCCs based on social media activity, the model provides a proactive layer of fraud detection that complements traditional systems. This dual layered approach improves detection accuracy, reduces false positives, and enhances the operational efficiency of fraud-detection workflows, thereby providing financial institutions with a robust, scalable, and forward-thinking solution to combat fraud.

REFERENCES:

[1] I. Mekterović, M. Karan, D. Pintar and L. Brkić, "Credit card fraud detection in card-not-present transactions: Where to invest?" Applied Sciences (Basel), 11(15), 2021. https://doi.org/10.3390/app11156766

[2] N. Akdemir and S. Yenal, "Card-not-present fraud victimization: A routine activities approach to understand the risk factors," Güvenlik Bilimleri Dergisi, 9(1), 2020. https://doi.org/10.28956/gbd.736179

[3] P. Chatterjee, D. Das and D. B. Rawat, "Digital twin for credit card fraud detection: Opportunities, challenges, and fraud detection advancements," Future Generation Computer Systems, 158, 2024. https://doi.org/10.1016/j.future.2024.04.057

[4] A. Abdallah, M. A. Maarof and A. Zainal, "Fraud detection system: A survey," Journal of Network and Computer Applications, 68, 2016. https://doi.org/10.1016/j.jnca.2016.04.007

[5] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi and S. Nerur, "Advances in social media research: Past, present and future," Information Systems Frontiers, 20(3), 2018. https://doi.org/10.1007/s10796-017-9810-y

[6] C. C. Yeh, Z. Zhuang, Y. Zheng, L. Wang, J. Wang and W. Zhang, "Merchant category identification using credit card transactions," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1736–1744.

[7] A. Razaque, M. B. H. Frej, G. Bektemyssova, F. Amsaad, M. Almiani, A. Alotaibi et al., "Credit card-not-present fraud detection and prevention using big data analytics algorithms," Applied Sciences (Basel), 13(1), 2022. https://doi.org/10.3390/app13010057

[8] V. Chandola, A. Banerjee and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys, 41(3), 2009. https://doi.org/10.1145/1541880.1541882

[9] H. Issa and M. A. Vasarhelyi, "Application of anomaly detection techniques to identify

fraudulent refunds," SSRN, 2011. https://doi.org/10.2139/ssrn.1910468

[10] W. Eberle and L. Holder, "Anomaly detection in data represented as graphs," Intelligent Data Analysis, 11(6), 2007. https://doi.org/10.3233/IDA-2007-11606

[11] S. Jayasingh and A. K. Swain, "Neural network in fraud detection," International Conference on Industrial Application of Soft Computing Techniques (IIASCT), 20, 2011, p. 22.

[12] S. Prasmaulida, "Financial statement fraud detection using perspective of fraud triangle adopted by SAS No. 99," Asia Pacific Fraud Journal, 1(2), 2016. https://doi.org/10.21532/apfj.001.16.01.02.24

[13] I. Ariyati, S. Rosyida, K. Ramanda, V. Riyanto, S. Faizah and Ridwansyah, "Optimization of the decision tree algorithm used particle swarm optimization in the selection of digital payments," Journal of Physics: Conference Series, 1641(1), 2020. https://doi.org/10.1088/1742-6596/1641/1/012090

[14] O. F. Prihono and P. K. Sari, "Comparison analysis of social influence marketing for mobile payment using support vector machine," Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 2019, pp. 367–374.

[15] L. Luyao, A. Al Mamun, N. Hayat, Q. Yang, M. E. Hoque and N. R. Zainol, "Predicting the intention to adopt wearable payment devices in China: The use of hybrid SEM-Neural network approach," PLoS ONE, 17(8), 2022. https://doi.org/10.1371/journal.pone.0273849

[16] A. Gupta, M. C. Lohani and M. Manchanda, "Financial fraud detection using naive bayes algorithm in highly imbalance data set," Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 2021. https://doi.org/10.1080/09720529.2021.1969733

[17] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," Expert Systems with Applications, 223, 2023. https://doi.org/10.1016/j.eswa.2023.119862

[18] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers and H. Mora, "Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining," Sustainability (Basel), 11(15), 2019. https://doi.org/10.3390/su11154235

[19] J. Y. Huang and J. H. Liu, "Using social media mining technology to improve stock price forecast accuracy," Journal of Forecasting, 39(1), 2020. https://doi.org/10.1002/for.2616

[20] M. Jaiswal and S. Das, "Detecting spam e-mails using stop word TF-IDF and stemming algorithm with Naïve Bayes classifier on the multicore GPU," International Journal of Electrical and Computer Engineering, 11(4), 2021. https://doi.org/10.11591/ijece.v11i4.pp3168-3175

[31] C. Chapman and K. T. Stolee, "Exploring regular expression usage and context in Python," Proceedings of the 25th International Symposium on Software Testing and Analysis, 2016, pp. 282–293. https://doi.org/10.1145/2931037.2931073

[32] H. Ali, M. N. Mohd Salleh, R. Saedudin, K. Hussain and M. F. Mushtaq, "Imbalance class problems in data mining: A review," Indonesian Journal of Electrical Engineering and Computer Science, 14(3), 2019. https://doi.org/10.11591/ijeecs.v14.i3.pp1552-1563

[33] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," Intelligent Data Analysis, 6(5), 2002. https://doi.org/10.3233/IDA-2002-6504

[34] C. H. Yu, "Resampling methods: Concepts, applications, and justification," Practical Assessment, Research, and Evaluation, 8(1), 2019.

[35] G. T. Sahid et al., "E-commerce merchant classification using website information," Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics, 2019, pp. 1–10.

[36] E. Rolf et al., "Representation matters: Assessing the importance of subgroup allocations in training data," International Conference on Machine Learning, 2021, pp. 9040–9051.

[37] A. C. Justice, K. E. Covinsky and J. A. Berlin, "Assessing the generalizability of prognostic information," Annals of Internal Medicine, 130(6), 1999. https://doi.org/10.7326/0003-4819-130-6-199903160-00016

[38] H. Zhai and X. Tong, "A practical Byzantine fault tolerance algorithms based on randomized mean clustering, trust and credibility," International Conference on Data Mining and Big Data, 2023, pp. 63–77.

[39] M. Abbas et al., "Multinomial naive bayes classification model for sentiment analysis,"

International Journal of Computer Science and Network Security, 19(3), 2019.

[40] T. Agrawal and T. Agrawal, Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient, Packt Publishing Ltd., 2021. https://doi.org/10.1007/978-1-4842-6579-6

[41] K. Jolly, Machine Learning with Scikit-learn Quick Start Guide: Classification, Regression, and Clustering Techniques in Python, Packt Publishing Ltd., 2018.

[42] D. Kartini, D. T. Nugrahadi and A. Farmadi, “Hyperparameter tuning using GridsearchCV on the comparison of the activation function of the ELM method to the classification of pneumonia in toddlers,” 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), 2021, pp. 390–395.

[43] D. Radulović and D. Negovanović, “Gait speed prediction based on walking parameters using MLPRegressor,” Ri-STEM, 13, 2021.

A. A. Farisi, Y. Sibaroni and S. Al Faraby, “Sentiment analysis on hotel reviews using multinomial naïve bayes classifier,” Journal of Physics: Conference Series, 1192(1), 2019. https://doi.org/10.1088/1742-6596/1192/1/012024

[44] S. E. Seide, K. Jensen and M. Kieser, “Utilizing radar graphs in the visualization of simulation and estimation results in network meta-analysis,” Research Synthesis Methods, 12(1), 2021. https://doi.org/10.1002/jrsm.1412

[45] M. Finck and F. Pallas, “They who must not be identified—Distinguishing personal from non-personal data under the GDPR,” International Data Privacy Law, 10(1), 2020. https://doi.org/10.1093/idpl/ipz026

[46] J. Soria-Comas and J. Domingo-Ferrer, “Big data privacy: Challenges to privacy principles and models,” Data Science and Engineering, 1(1), 2016. https://doi.org/10.1007/s41019-015-0001-x

[47] A. K. Schnackenberg, E. Tomlinson and C. Coen, “The dimensional structure of transparency: A construct validation of transparency as disclosure, clarity, and accuracy in organizations,” Human Relations, 74(10), 2021. https://doi.org/10.1177/0018726720933317

[48] A. Smith-Renner et al., “Digging into user control: Perceptions of adherence and instability in transparent models,” Proceedings of the 25th International Conference on Intelligent User Interfaces, 2020, pp. 519–530. https://doi.org/10.1145/3377325.3377491

[49] J. W. Gichoya, K. Thomas, L. A. Celi, N. Safdar, I. Banerjee, J. D. Banja et al., “AI pitfalls and what not to do: Mitigating bias in AI,” British Journal of Radiology, 96(1150), 2023. https://doi.org/10.1259/bjr.20230023

[50] D. C. Mohr, K. Cheung, S. M. Schueller, C. Hendricks Brown and N. Duan, “Continuous evaluation of evolving behavioral intervention technologies,” American Journal of Preventive Medicine, 45(4), 2013. https://doi.org/10.1016/j.amepre.2013.06.006

[51] S. Sadiq and G. Governatori, “Managing regulatory compliance in business processes,” in Handbook on Business Process Management 2: Strategic Alignment, Governance, People and Culture, Springer Berlin Heidelberg, 2014, pp. 265–288.

[52] J. Drydyk, “Accountability in development: From aid effectiveness to development ethics,” Journal of Global Ethics, 15(2), 2019. https://doi.org/10.1080/17449626.2019.1636116

[53] J. R. Mitchell, R. K. Mitchell, R. A. Hunt, D. M. Townsend and J. H. Lee, “Stakeholder engagement, knowledge problems and ethical challenges,” Journal of Business Ethics, 175(1), 2022. https://doi.org/10.1007/s10551-020-04550-0

[54] A. C. Ozioko, “The use of artificial intelligence in detecting financial fraud: Legal and ethical considerations,” Multi-Disciplinary Research and Development Journals International, 5(1), 2024.

[55] C. Petrozzino, “Big data analytics: Ethical considerations make a difference,” Scitech Lawyer, 16(3), 2020.

[56] M. Ait Said and A. Hajami, “AI methods used for real-time clean fraud detection in instant payment,” International Conference on Soft Computing and Pattern Recognition, Springer International Publishing, 2021, pp. 249–257.

[57] M. Ait Said and A. Hajami, “Card-Not-Present fraud detection: Merchant category code prediction of the next purchase,” International Conference on Intelligent Systems Design and Applications, Springer Nature Switzerland, 2022, pp. 92–98.

[58] M. Ait Said, A. Hajami and A. Krari, “Behavioral profiling for card-not-present fraud detection leveraging ISO8583 data to identify anomalous patterns,” Journal of Theoretical and Applied Information Technology, 103(5), 2025.