Sentiment Analysis of Case Suspects In Digital Forensics and Legal Analytics

Sundar Krishnan *skrishnan@shsu.edu*
Department of Computer Science
Sam Houston State University
Huntsville, TX, USA

Narasimha Shashidhar *karpoor@shsu.edu*
Department of Computer Science
Sam Houston State University
Huntsville, TX, USA

Cihan Varol *cxv007@shsu.edu*
Department of Computer Science
Sam Houston State University
Huntsville, TX, USA

ABM Rezbaul Islam *ari014@shsu.edu*
Department of Computer Science
Sam Houston State University
Huntsville, TX, USA

Abstract

Sentiments of suspects in a legal case or digital forensic investigation can be of use when profiling their state of mind or feelings. Such information can help case investigators to plot their actions against case timelines, understand their instincts and build psychological profiles. In this research, the authors first assemble a fictional dataset of electronic evidence and store in a SQL Database. Next, they leverage different analytical techniques, automation, and Natural Language Processing (NLP) to propose an approach to plot sentiments of case. This information is then presented via a custom software for case investigators who can use it to pick a suspect from the case and obtain their sentiments against various electronic evidence sources within the case load that they were associated with.

Keywords: Sentiment Analysis, Suspect Profiling, Legal Case Evidence, Electronic Stored Information, eDiscovery, Digital Forensic Analytics, Digital Forensics, Supervised Learning, Unsupervised learning, Neural Networks, Legal Analytics, Machine Learning, Natural Language Processing.

1. INTRODUCTION

Human sentiment can be defined as an attitude, thought, emotion, opinion, or judgment prompted by feeling intended to be conveyed by words, acts, or gestures (Merriam-Webster, n.d.), (Dictionary.com, n.d.). Sentiment analysis is often undertaken through natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information ("Sentiment analysis - Wikipedia," n.d.). With the recent advances in deep learning and ever-increasing volumes of data, the ability of algorithms to analyze text for sentiments has greatly improved. Legal cases and digital forensic investigations often deal with vast amounts of digital data related to the case or digital forensic investigation. Employing analytic techniques such as Neural Networks (NN), Machine Learning (ML), and Artificial Intelligence (AI) can greatly assist in mining sentiments of suspects or actors from within the digital data, thus reducing costs, rework and time while increasing review process efficiency. Practical use of sentiment analysis is in suspect profiling during a case investigation. Criminal or suspect profiling can be defined and described as a technique whereby the probable characteristics of the suspect/offender are predicted based on the behaviors exhibited in the commission of a crime (Kocsis, 2006). This technique can also be an investigative tool wherein skilled profilers look at the crime from the behavioral aspect. Few inputs to such profiling are suspect's behavior, gender, age, geography, personality, sentiment, and psychology. Often experienced and skilled FBI profilers are engaged for in-depth suspect profiling. However, low-level suspect profiling is always undertaken by forensic and legal professionals. Recent sentiment analysis has been popularized by leveraging machine learning algorithms wherein digital data in the form of text is analyzed for sentiment polarity, from positive to negative. The learning of such algorithms to predict or classify sentiments is undertaken by supervised or unsupervised learning. In a supervised setting, training the machine learning algorithms with examples of emotions in texts can assist the machine to automatically learn how to detect sentiments without human input. Sentiment analysis algorithms often use Natural Language Processing (NLP) techniques to mine the text for linguistics and, at the minimum, predict if it was positive or negative. Such techniques can be used to repeatedly predict the sentiment as each token in a piece of ingested text.

In the legal world, electronic discovery (eDiscovery) is often the most expensive component of litigation, given the growth of technology in our lives and in business. Levering technology advances such as Technology Assisted Review(TAR), ML, NN, and AI has been demonstrated in the eDiscovery industry as a huge cost-savings benefit, and these savings can be substantial to a case (Casepoint, 2020), ("How Artificial Intelligence Saves Money on eDiscovery | IDISCOVER Global," 2018), ("How Legal Teams Can Leverage Technology to Minimize eDiscovery Costs," 2018). Ina digital forensic setting, forensic tools have embarked on implementing machine learning and deep learning for face recognition, pattern recognition in images, etc., within processing of digital forensic evidence. Such assistance for a forensic investigator can help in speedier forensic analysis, reduce costs and time (Ron J, n.d.). Similarly, eDiscovery tools have started to leverage ML and AI("How AI is transforming eDiscovery industry | Casepoint," n.d.), ("3 Uses of AI in ediscovery | EDRM," n.d.). As legal cases and forensic investigations can often deal with enormous digital data (big data), sentiment mapping of suspects or actors can be useful in filtering down to relevant case evidence (scope) for deeper analysis. Few other uses of sentiment analysis is suspect (offender) profiling, personality, mapping their attitude and moods. Such information can give a better insight into the suspect and narrow it down their communication and activities for legal arguments. In a forensic setting and early case assessments, such information can help forensically profile suspects and assist with their psychological studies. In this paper, the authors employ sentiment analysis against three fictitious legal caseloads using various NLP and Neural network supervised learning techniques. The caseloads are randomly assembled from various real-life public potential sources of evidence. The authors demonstrate that the sentiments of suspects/actors can be mapped against timelines of the case and thus assist in quicker case review and digital forensic analysis. Our results are presented on a custom application that can be used by the case investigator to narrow down suspect's sentiments against case timelines.

2. LITERATURE REVIEW

Sentiment analysis through data mining and artificial intelligence has been well researched to extract, and predict subjective information, such as opinions, attitudes, emotions, and feelings. With the growth of smartphone and their involvement in many criminal cases, they are a valuable source of suspect profiling. Zhang et al. (L. Zhang, Hua, Wang, Qian, & Zheng, 2014)analyzed mobile reviews and could classify emotions based on certain characteristics. Sentiment analysis of text data from the Internet has long been a popular research topic. Hussain et al. (Hussain, Kanakam, Suryanarayana, & Gupta, 2017) analyze web session logs to the behavioral pattern of a user from web browsing logs. Sentiment analysis has also been viewed as a text classification problem that needs labeled data to train the model. Salahet al. (Salah & Gayar, 2019) used the Enron emails as training data and built a classifier using a supervised machine learning algorithm for email labeling. Large amounts of texts are often found in emails. Mcguire et al. (Mcguire & Leung, 2018)proposed an approach and applied sentiment analysis to a system prototype for investigators to identify points of interest in emails according to the attitude of authors with respect to the content of their emails. Bogawar et al.(Bogawar & Bhoyar, 2016) apply data mining approaches such as k-means clustering, fuzzy c-means clustering, and neural network back propagation algorithm for classification of emails as per the sentiments within them. They compare the analysis of these and find the neural network back propagation algorithm giving the best sentiment recognition rate. Twitter data has often attracted research interests due to its popularity with world leaders and movie stars. Wang et al. (Wang, Kim, Lee, & Youn, 2018) propose a modified ChiSquare-based feature clustering and weighting scheme for the sentiment analysis of Twitter messages that outperforms existing analysis schemes irrespective of the size of training and test data. They consider parts of speech tagging, discriminatory words, dependency of the words, and influence of emotional words. Within a computer disk, forensic investigators often spend a vast amount of time when interpreting evidence along case timelines. Studiawan et al. (Studiawan, Sohel, & Payne, 2020)propose a sentiment analysis technique for forensic investigators to automatically extract events of interest from log messages in the forensic timeline. They apply it on four public forensic cases with high accuracy. Text can also be misused for harassment highlighting negative sentiments and false information. Budiman et al. (Budiman, Zaatsiyah, Niswah, & Faizi, 2020)used Naïve Bayes method to identify sexual harassment in Indonesian Twitter data. They find that negative sentiments are higher than positive sentiments and conclude that Twitter can be a medium for verbal sexual harassment. Sivasangariet al.(Sivasangari, Mohan, Suthendran, & Sethumadhavan, 2018) use VADER based classifier for sentiment analysis to identify false facts in tweets. Text can contain Emojis, abbreviations, slang, and special characters. Shiha et al. (Shiha & Ayvaz, n.d.)study the impact of Emojis on sentiments using Twitter data and conclude that Emoji characters appear to have a higher impact on overall sentiments of the positive opinions when compared to negative opinions. Text can also be in multiple languages making sentiment analysis a bit complicated. Al-Rowaily et al. (Al-Rowaily, Abulaish, Al-Hasan Haldar, & Al-Rubaian, 2015) have developed a Bilingual Sentiment Analysis Lexicon consisting of English and Arabic Lexicon for opinion mining and sentiment analysis. Al-Smadi et al. (Al-Smadi, Qawasmeh, Al-Ayyoub, Jararweh, & Gupta, 2018)trained and implemented two approaches of deep recurrent neural network (RNN) and support vector machine (SVM) on reviews of Arabic Hotels' to highlight the challenges of aspect based sentiment analysis in the Arabic language. Shalini et al.(Shalini et al., 2018) used Convolutional Neural Network (CNN) to classify Bengali-English code mixed data as positive, negative or neutral. A lot about humans can be gathered from the Internet and their smart device usage. Stachl et al. (Stachl et al., 2020) predict personality from patterns of behavior collected through smartphone usage and conclude that communication and social behavior is the most predictive among other personality dimensions. Cheng et al. (Cheng, Chandramouli, & Subbalakshmi, 2011) investigate text gathered from social networking applications on the Internet to predict the gender of authors. They also propose multiple psycho-linguistic and gender-preferential cues to build the feature space for gender identification. Sentiments also help narrow down gender of the person. Rafi (Rafi, 2008) analyzed SMS messages along with perceptions of males and females and found that SMS language was influencing language of media and a significant difference can be found between male and female texters' linguistic properties. Ahmad et al.(Khdr & Varol, 2019) predicted age and gender from SMS text messages using Naïve Bayes, Support Vector Machine, and J48 decision trees.

Offender/Suspect or criminal profiling and cyber forensics also involve human sentiment and behavior analysis. Orebaugh et al (Orebaugh & Allnutt, 2009) present a framework for Instant Messaging(IM) authorship analysis and a feature set taxonomy for use in cyber forensics and cybercrime investigations. Fuzail et al. (Ahmad, Syal, & Tinna, 2018) propose a criminal policing approach by applying Rossmo's equation using local crime sentiments. While sentiment of suspect in a criminal case can be a low-cost exercise, given the narrow scope of the investigation and warrants, civil litigation can be just the opposite. In civil legal cases, the digital forensic investigation and caseload analysis can be expensive and time consuming given the scope of the exercise, multiple human suspects/players in the investigation scope, and the extensive use of smart devices (potential evidence) in our lives. Krishnan et al. (Krishnan & Shashidhar, 2021), (Krishnan, 2019) discuss the costs of legal costs during an investigation. While there are plenty of academic articles in sentiment analysis and is a well-researched topic, little to no research has been undertaken for suspect's sentiments in an investigation/eDiscovery setting. In this research, the authors propose utilizing machine learning and neural network approaches to work with vast amounts of data that is typically encountered in case evidence. Automating the various algorithms in supervised learning setting and allowing the user to provide feedback from a custom tool allows for tuning training data and thus improving model accuracy over time.

3. METHODOLOGY

As real-life forensic investigation evidence or a legal caseload is usually not readily available in public or for academic research, the authors felt the need to build custom caseloads of electronic evidence/Electronic Stored Information (ESI) and later make them available for academic research.

3.1 Experiment Design

The research experiment revolved around three fictitious legal case ESI or digital forensic investigation evidence (datasets). The experiment was carried out using an Intel(R) Core(TM) i5-3470 CPU @ 3.20GHz 16 GB RAM PC and a 64-bit Windows 10 operating system. Each caseload ESI contained emails (public sources and custom) (Technologies, n.d.-b), (Technologies, n.d.-a),SMS and publicly available WhatsApp data ("UCI Machine Learning Repository: SMS Spam Collection Data Set," n.d.), publicly available Twitter data ("Twitter US Airline Sentiment," 2019), (Agarwal, 2018), ("First GOP Debate Twitter Sentiment," 2018),publicly available Facebook data(Bencina, 2017), (K. Zhang et al., 2011) and a few custom random MS Word files. Each case ESI was further updated to introduce random suspect names and few posts and tweets were updated to highlight a case/investigation scenario. The labeled dataset for supervised learning was obtained from movie reviews specially selected for sentiment analysis containing 25,000 samples of reviews with binary sentiments ("Bag of Words Meets Bags of Popcorn, Use Google's Word2Vec for movie reviews," 2014).Figure 1 highlights the overall design of the experiment. Given a suspect/POI from the case/investigation, the user (investigator)

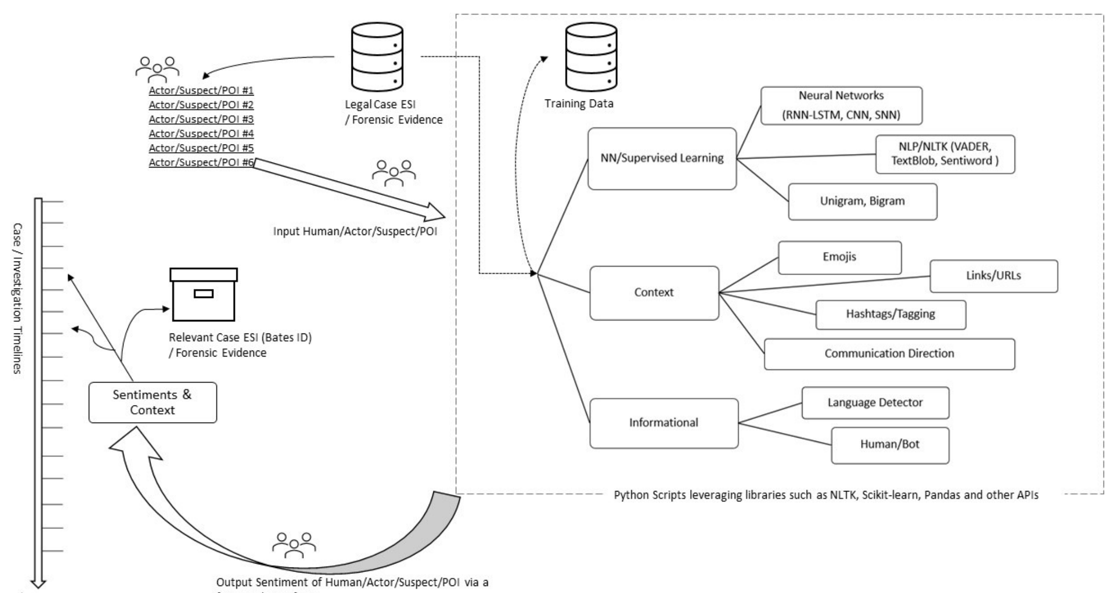

FIGURE 1: Sentiment Analysis of a case suspect or Person of Interest (POI) within the forensic investigation timeline or case eDiscovery scope.

of the tool can obtain the sentiment expressed by him/her, narrow down to the document/evidence, and obtain the timelines. The tool is automated to fetch from the database the sentiment results and other data for the investigator. The sentiments and data normalization can also be re-run/executed on demand from the tool.

3.2 Dataset - Preparation and Normalization

Each case ESI dataset was first assembled as flat-files from various public sources. These flat files were then ingested using custom C#.NET programs into three SQL Server databases running on a local SQL Server instance. A .pst parser "Pstxy" ("Pstxy: Outlook .pst and .ost file reader for .Net," n.d.) was used to parse emails. SQL tables were created for each type of data being injected. For training data, a separate database was created on the same instance. For each flat-file being ingested, file metadata was also identified and uploaded into the table. A new key column for bates-id (document id) was introduced to represent each ingested file. This resulted in three databases housing three different case ESI or digital forensic evidence for investigation. For each case ESI (databases), names of people (from the flat files) were randomly changed for fictitious actors/suspects of the case. Timestamp was randomly updated to reflect case/investigation timelines. Data for retweets, case timelines, followers, likes, the direction of communication and reaction was randomly added to each post/tweet. In few instances, Tweets and Facebook posts were modified to build the scenario conversations for each case/investigation. Care was taken to randomize at every opportunity to avoid bias. For each type of case/investigation ESI (evidence) that is now on database tables, text data was normalized into separate tables. Sentence identification methods were employed to parse large documents and paragraphs. Each sentence was stored as a row in the normalized tables. Emojis, abbreviations, glyphs, grapheme clusters, email addresses, social media identifiers or handles, hash tags, acronyms, and URLs were extracted into separate columns and stored in the normalized tables. This was done as input text for analytics can sometimes consist of garbage/Html text, especially when dealing with Unicode (UTF-8 encoded) and umlauts. Also, such information, when extracted, can help build a context for the investigator or legal mind using the proposed analytics tool.

3.3 Sentiment Analysis using Supervised and Hybrid Learning

The IMDB movie reviews sentiment dataset ("Bag of Words Meets Bags of Popcorn, Use Google's Word2Vec for movie reviews," 2014) was used for supervised learning. Feature Selection was not undertaken for analytics as the only feature to focus on was the text from the normalized tables. Different approaches outlined below were taken for sentiment analysis using Python. Database connection was established to first train the model and then apply it against the normalized data. The reason to use different approaches was to provide the user with choices and seek feedback on correct/incorrect sentiment categorization, thereby update the training data in a loop. The below NLP/NLTK and Neural Networks techniques were used. Randomization during labeled data selection for training and testing was avoided to allow for uniformity in results for the user of the custom tool, lest each run of analytical algorithms against case evidence would end with confusing different results confusing the user. The various approaches used are listed below.

1. Applied VADER (Valence Aware Dictionary and Sentiment Reasoner) (Hutto & Gilbert, 2014) to each of the normalized case datasets in a supervised learning model using training and testing data in a 80-20 ratio (20000 for training and 5000 for testing) with no randomization. VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to feelings and sentiments expressed in social media by considering individual tokens for sentiment analysis. VADER was installed using the command "pip install vaderSentiment" at the terminal window. Sentiment polarity against each sentence and accuracy of the model was calculated and stored in a separate database table. The Sentiment Intensity Analyzer class methods provided a sentiment intensity score to each text sentence.

2. Applied SentiWordNet (Esuli & Sebastiani, 2006) (an opinion lexicon derived from the WordNet (“WordNet | A Lexical Database for English,” 2010) database) against each of the normalized case datasets in a supervised learning model using training and testing data in a 80-20 ratio (20000 top samples for training and 5000 bottom samples for testing) with no randomization. SentiWordNet is a lexical resource for opinion mining and is publicly available for research purposes. Python’s NLTK provides both SentiWordNet and WordNet classes for import. SentiWordNet approach computes the polarity of the words and averages the value. Sentiment polarity against each sentence and accuracy of the model was calculated and stored in a database table for each of the datasets.

3. Applied Text Blob(“TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation,” n.d.) against each of the normalized case datasets in a supervised learning model using training and testing data in a 80-20 ratio (20000 top samples for training and 5000 bottom samples for testing) with no randomization. TextBlob is a Python (2 and 3) library for processing textual data. TextBlob is a python library offering a simple API to access its methods and perform basic NLP tasks. Naïve Bayes was used as the classifier from Textblob library of classifiers as it offered better accuracy compared to MaxEntClassifier. Sentiment polarity against each sentence and accuracy of the model was calculated and stored in a database table for each dataset.

4. Applied Unigram approach along with Python’s NLTK Sentiment Analyzer against each of the normalized case datasets in a supervised learning model using training and testing data. Unigrams or 1-gram is an N-gram with simply one string in a text. A Naïve Bayes classifier was used. Due to memory limitations, the training set was limited to top 3000samples, and the testing set was bottom 600 samples (out of a total of 25,000 in the labeled dataset) with no randomization. Sentiment polarity against each sentence and accuracy of the model was calculated and stored in a database table for each dataset.

5. Applied Bigram approach against the normalized case data in a supervised learning model using training and testing data. Bigram or 2-gram is an N-gram that is typically a combination of two strings or words that appear in a text. A Naïve Bayes classifier was used. Due to memory limitations, the training set was limited to top 5000 samples, and the testing set was bottom 1000 samples (out of a total of 25,000 in the labeled dataset). Sentiment polarity against each sentence and accuracy of the model was calculated and stored in database table for each dataset.

6. Applied a Long short-term memory (LSTM) (Recurrent Neural Network) against each of the normalized case datasets in a supervised learning model using training and testing data. Trained on 16000 samples and validated on 4000 samples from the labeled dataset with 6 Epochs, 128 neurons and no randomization. Sentiment polarity against each sentence and accuracy of the model was calculated and stored in a database table for each dataset.

7. Applied a Convolutional Neural Network (CNN) against each of the normalized case datasets in a supervised learning model, using training and testing data. Trained the model on 16000 samples and validated on 4000 samples from the labeled dataset. Used 6 epochs using GloVe embeddings (Pennington, Socher, & Manning, 2014) to create our feature matrix, one dimensional Convolutional layer with 128 features, kernel size of 5 and activation function as relu. Finally, we added a dense layer and used activation function as Sigmoid. Sentiment polarity against each sentence and accuracy of the model was calculated and stored in a database table for each dataset.

8. Applied a Simple Neural Network against each of the normalized case datasets in a supervised sequential learning model using 16000 samples for training and 4000 samples as testing data, embedding layer of 100, final dense layer with activation function as Sigmoid, and 10 epochs. Used using GloVe embeddings to create our feature

matrix. Sentiment polarity against each sentence and accuracy of the model was calculated and stored in a database table for each dataset.

3.4 Presentation
A custom Windows Forms application screen was designed and developed to help steer the case investigator to use the sentiment analysis of a case suspect. MS-SQL stored procedures and queries were written to automate the display and enrich context by leveraging data about URLs, emojis, timelines, etc. A simple logic was added to determine if the suspect was a bot.

4. ANALYSIS
Upon application of sentiment analysis algorithms, and upload of the case dataset, the investigator can now pick a suspect/POI from the case to obtain his/her sentiments as classified by the different models. Each analytical model yielded sentiment results and accuracy based on size of training and test sets. ROC and AUC were calculated for cross verification.

1. VADER Using Python's VADER library, the whole labeled dataset was used in an 80-20 ratio for supervised learning (top 20,000 samples for training and bottom 5000 for testing).A model accuracy of 69.1% was achieved. Figure 3displays the model's ROC curve with AUC = 0.69.

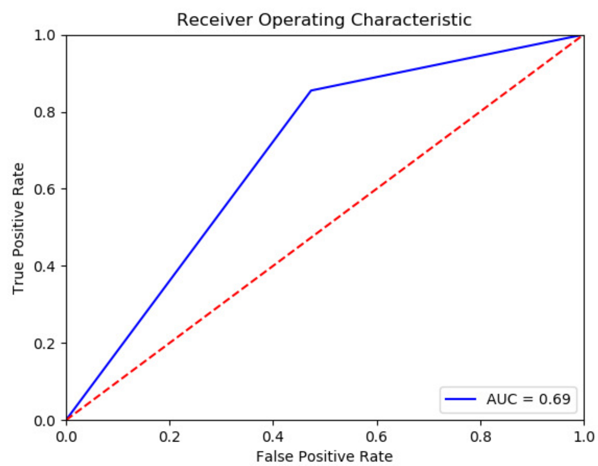

FIGURE 3: ROC - VADER (Valence Aware Dictionary and Sentiment Reasoner).

2. Using Sentiwordnet from Python's NLTK library, the whole labeled dataset was used in an 80-20 ratio for supervised learning (top 20,000 samples for training and bottom 5000 for testing). A model accuracy of 66.7% was achieved. Figure 4 displays the model's ROC curve with AUC= 0.67.

3. TextBlob Using Textblob approach and Naïve Bayes Classifier, top 2000 samples were used for training and bottom 400 for testing due to memory limitations. A classifier accuracy of 84.3% was achieved. Figure 5 displays the model's ROC curve with AUC = 0.68.

4. Using Unigram approach, NLTK Sentiment Analyzer, and Naïve Bayes Classifier, a model accuracy of 78.2% was achieved. Due to memory limitations, small sizes of training and test data was used. However, we noticed that increasing the training set lowered the accuracy. Figure 6 displays the ROC curve with AUC = 0.5. However, the Precision is high (0.72), Recall is high (0.85) and F-measure is high (0.78). High accuracy but a low AUC value implies that the training features may be imbalanced i.e., there are much more negative sentiments than positives taken in consideration during training. The

apparent discrepancy has to do with the lack models' success at identifying true negatives.

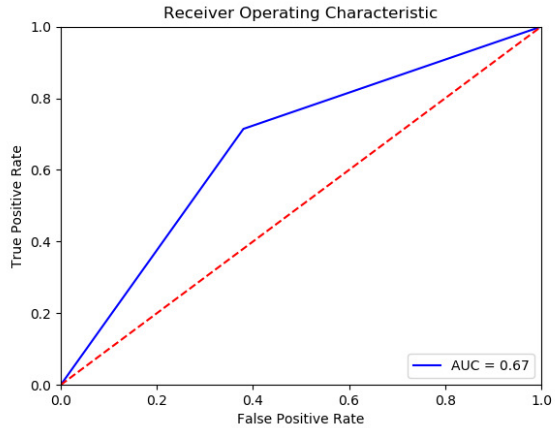

FIGURE 4: ROC – Sentiwordnet.

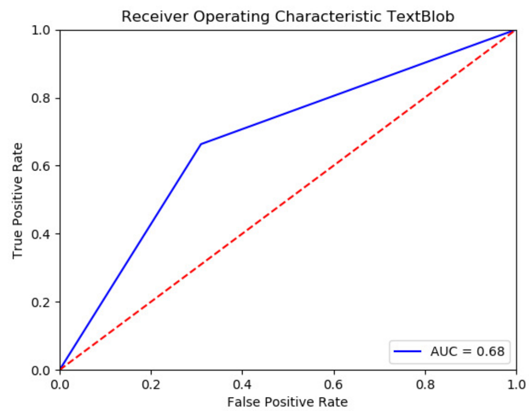

FIGURE 5: ROC – TextBlob.

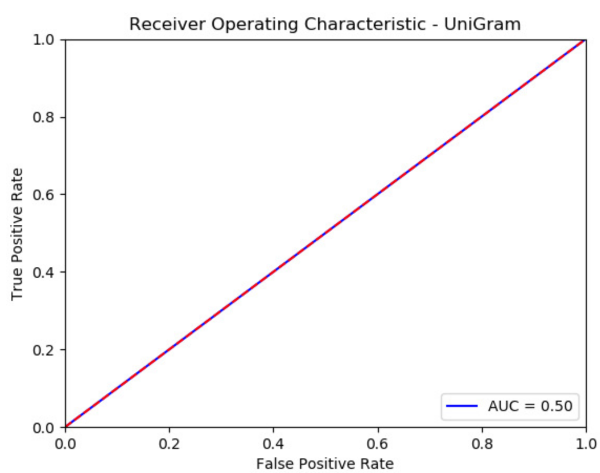

FIGURE 6: ROC – Unigram.

5. Using the Bigram approach increasing training set size decreased the accuracy and caused memory issues. Thus, a subset of the labeled dataset considered, and model accuracy achieved was 54.6%. Figure 7 displays the ROC curve with AUC at 0.5. However, the Precision was 0.54, recall was low at 0.02 and F-measure was 0.042.

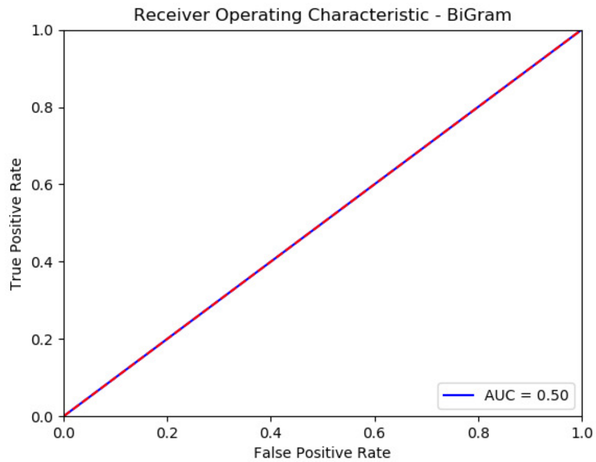

FIGURE 7: ROC –Bigram.

6. Using the RNN approach, to compile our model, we used the adam optimizer, binary cross entropy as our loss function and accuracy as metrics. We started with initializing a sequential model followed by the creation of the embedding layer. Next, we created a LSTM layer with 128 neurons. We used GloVe embeddings to create our feature matrix. The model's accuracy was 80.6%. Figure 8 displays the ROC curve with AUC at 0.92.

7. Using the CNN approach, to compile our model, we used the adam optimizer, binary cross entropy as our loss function and accuracy as metrics. We used GloVe embeddings to create our feature matrix. The model's accuracy was 84.1%. Figure 9 displays the ROC curve with AUC at 0.92.

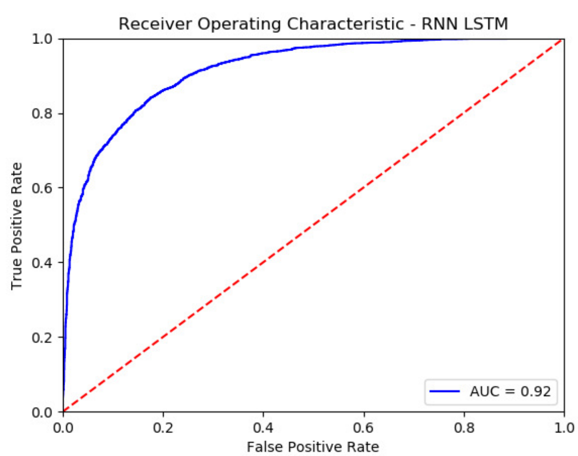

FIGURE 8: ROC – RNN.

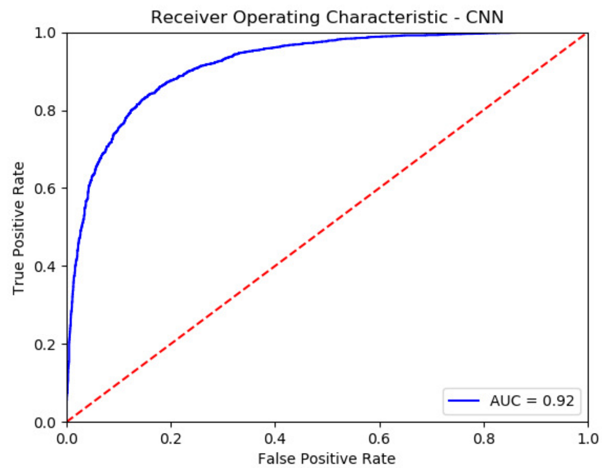

FIGURE 9: ROC – CNN.

8. Using the Simple Neural Network approach, to compile our model, we used the adam optimizer, binary cross entropy as our loss function and accuracy as metrics. We used GloVe embedding to create our feature matrix. The model's accuracy was 69.4%.Figure 10 displays the ROC curve with AUC at 0.76.

Table 1 summarizes the results all the models/approaches. Figure 10 displays the custom Windows Forms application that was developed to help steer the case investigator to use the sentiment analysis against case suspects/POI. At run-time, all suspects/POI in the case dataset (ESI) was listed on the screen. Upon the user picking one of the suspects/POI, sentiments from various algorithms listed above were displayed along with accuracy. Database stored procedures and queries were executed by the custom application to automate the display of information on the screen.

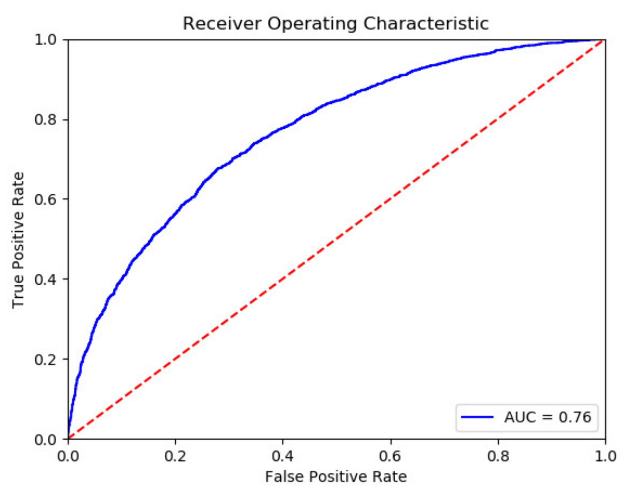

FIGURE 9: ROC– Simple Neural Network.

Model/Approach	AUC	Accuracy (%)
CNN	0.92	81.4
LSTM (RNN)	0.92	80.6
SentiWordNet	0.67	66.7
Simple Neural Network	0.76	69.4
Unigram	0.50	78.2
Bigram	0.50	54.6
TextBlob	0.68	84.3
VADER	0.69	69.1

TABLE 1: Summary of results.

A logic for bot detection for tweets/posts was also incorporated on the tool. To detect bot activity, few indicators like - high volumes of activity, a high percentage of retweets, many followers with less followed, etc. were added to custom software program. Average sentiment on sentences per document (bates id) per person was then displayed per algorithm. Detailed sentiments of the person per document was also presented on the screen. Other details such as timelines and context were also displayed. This would be helpful to the case investigator as background details of the suspect/POI are sometimes instrumental in building the suspect's psychological profile. The tool also allowed for the case investigator to manually mark a sentiment as correct or incorrect and this was fed back into the labeled training dataset to be reused in the next analysis run(supervised learning) along with the existing labeled dataset. The tool also allowed for the case investigator to pick specific date range within the case timelines as a filter. All the code files for this research are available on GitHub (Krishnan, n.d.)

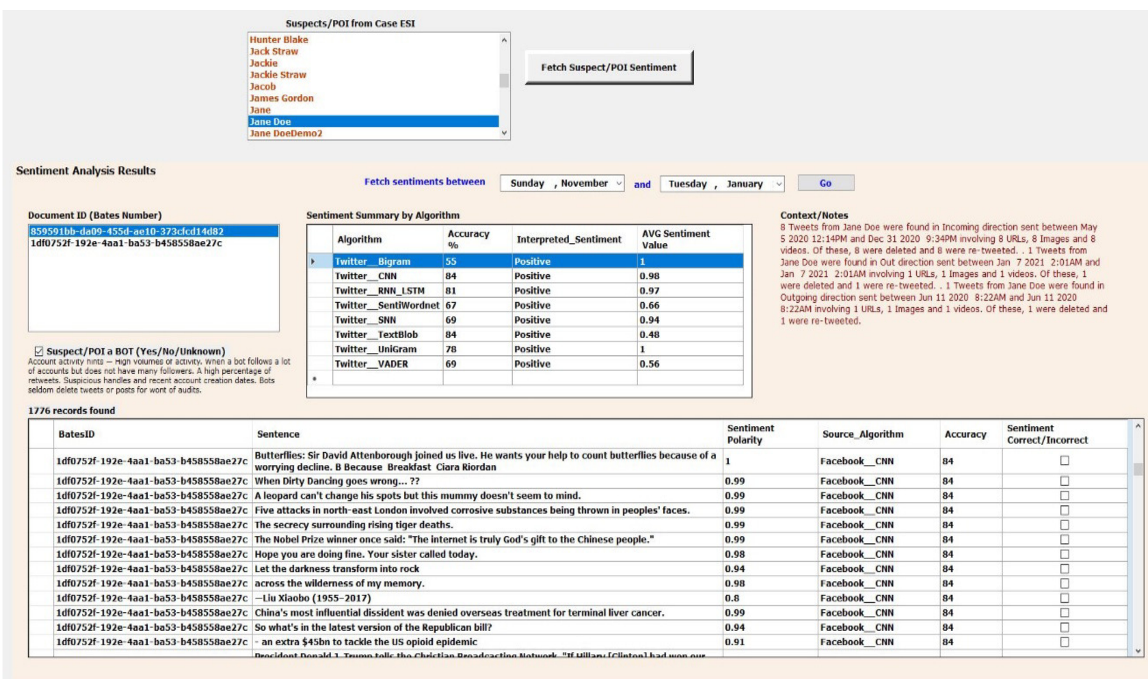

FIGURE 10: Custom software (tool) for sentiment analysis.

5. CONCLUSION

Leveraging machine learning into legal analytics and digital forensics can be of immense help as electronic case artifacts have grown in volume and disparate sources. A typical case to investigate may involve processing large amounts of electronic data in the quest for something such as sentiments expressed by people involved in the case. There may be a few suspects in the case that the investigator may want to focus on. In this research, the authors demonstrate the use of various machine learning and neural network approaches to process legal/forensic case evidence (ESI) and mine sentiments of suspects involved in the case. Fictitious (synthetic) case datasets were assembled from custom and public sources, and various analytical approaches for sentiments coupled with a custom tool was developed. In addition to displaying a comparative viewpoint, the use of multiple analytical approaches allows the investigator to pick a particular approach over the other and purse their investigation. This avoids bias in analytical technique selection from the very beginning. The custom software allows for fine-tuning the training dataset over time due to a user feedback loop, thereby allowing for improved model accuracy over time and use. Thus, the software helps reduce analysis time, reduces costs of the case investigator to analyze electronic data from the case pile for suspect sentiments and reduces rework effort. Data cleansing techniques(preprocessing) employed on case ESI, and the quality of the training dataset used can greatly affect the overall results of the various analytical models. Overall, this proposed approach gives insights into suspects of the case to retain or eliminate them during an investigation. In future work, the authors plan to incorporate additional training datasets and expand support for other languages.

6. REFERENCES

Agarwal, R. (2018). Twitter hate speech. Retrieved February 14, 2021, from https://www.kaggle.com/vkrahul/twitter-hate-speech.

Ahmad, F., Syal, S., & Tinna, M. S. (2018). Criminal Policing Using Rossmo's Equation by Applying Local Crime Sentiment. *Advances in Intelligent Systems and Computing*, *542*, 627–637. https://doi.org/10.1007/978-981-10-3223-3_60

Al-Rowaily, K., Abulaish, M., Al-Hasan Haldar, N., & Al-Rubaian, M. (2015). BiSAL – A bilingual sentiment analysis lexicon to analyze Dark Web forums for cyber security. *Digital Investigation*, *14*, 53–62. https://doi.org/10.1016/J.DIIN.2015.07.006.

Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., & Gupta, B. (2018). Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *Journal of Computational Science*, *27*, 386–393. https://doi.org/10.1016/J.JOCS.2017.11.006.

Bag of Words Meets Bags of Popcorn, Use Google's Word2Vec for movie reviews. (2014). Retrieved October 25, 2021, from https://www.kaggle.com/c/word2vec-nlp-tutorial/data.

Bencina, J. (2017). Facebook News Scraper. Retrieved February 14, 2021, from https://github.com/jbencina/facebook-news.

Bogawar, P. S., & Bhoyar, K. K. (2016). Soft computing approaches to classification of emails for sentiment analysis. *ACM International Conference Proceeding Series*, *25-26-Augu*. https://doi.org/10.1145/2980258.2980304.

Budiman, K., Zaatsiyah, N., Niswah, U., & Faizi, F. M. N. (2020). Analysis of Sexual Harassment Tweet Sentiment on Twitter in Indonesia using NaÃ¯ve Bayes Method through National Institute of Standard and Technology Digital Forensic Acquisition Approach. *Journal of Advances in Information Systems and Technology*, *2*(2), 21–30. Retrieved from https://journal.unnes.ac.id/sju/index.php/jaist/article/view/44305.

Casepoint. (2020). *Case Study: How a Multi-Billion Dollar Corporation Reaped Major Cost*

Savings in an Internal Investigation Leveraging Casepoint Case Study: How a Multi-Billion Dollar Corporation Reaped Major Cost Savings in an Internal Investigation Leveraging Casepoint C.

Cheng, N., Chandramouli, R., & Subbalakshmi, K. P. (2011). Author gender identification from text. *Digital Investigation*, *8*(1), 78–88. https://doi.org/10.1016/J.DIIN.2011.04.002.

Dictionary.com. (n.d.). Sentiment Definition & Meaning | Dictionary.com. Retrieved August 3, 2021, from https://www.dictionary.com/browse/sentiment.

Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*. European Language Resources Association (ELRA). Retrieved from http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.

First GOP Debate Twitter Sentiment. (2018). Retrieved February 14, 2021, from https://www.kaggle.com/crowdflower/first-gop-debate-twitter-sentiment.

How AI is transforming eDiscovery industry | Casepoint. (n.d.). Retrieved October 24, 2021, from https://www.casepoint.com/blog/ai-transforming-ediscovery/.

How Artificial Intelligence Saves Money on eDiscovery | IDISCOVER Global. (2018). Retrieved August 5, 2021, from https://idiscoverglobal.com/how-artificial-intelligence-saves-money-on-ediscovery/.

How Legal Teams Can Leverage Technology to Minimize eDiscovery Costs. (2018). Retrieved August 5, 2021, from https://www.jndla.com/blog/how-legal-teams-can-leverage-technology-minimize-ediscovery-costs.

Hussain, S. M., Kanakam, P., Suryanarayana, D., & Gupta, S. (2017). Forensics Data Analysis for Behavioral Pattern with Cognitive Predictive Task. *Communications in Computer and Information Science*, *827*, 549–557. https://doi.org/10.1007/978-981-10-8657-1_41.

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

Khdr, A. J., & Varol, C. (2019). Age and Gender Identification by SMS Text Messages. *2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018*. https://doi.org/10.1109/IDAP.2018.8620780.

Kocsis, R. N. (2006). What Is Criminal Profiling? In *Criminal Profiling* (pp. 1–11). Humana Press. https://doi.org/10.1007/978-1-59745-109-3_1.

Krishnan, S. (n.d.). Project · GitHub. Retrieved May 6, 2022, from https://github.com/kshsus.

Krishnan, S. (2019). eDiscovery Challenges in Healthcare. *International Journal of Information Security Science*, *8*(2), 30–43.

Krishnan, S., & Shashidhar, N. (2021). Interplay of Digital Forensics in eDiscovery. *International Journal of Computer Science and Security (IJCSS)*, *15*(2), 19. Retrieved from https://www.cscjournals.org/library/manuscriptinfo.php?mc=IJCSS-1602.

Mcguire, J. C., & Leung, W. S. (2018). Enhancing Digital Forensic Investigations Into Emails Through Sentiment Analysis. In *European Conference on Cyber Warfare and Security;*

Merriam-Webster. (n.d.). Definition of Sentiment by Merriam-Webster. Retrieved August 3, 2021, from https://www.merriam-webster.com/dictionary/sentiment#synonym-discussion.

Orebaugh, A., & Allnutt, D. J. (2009). Data Mining Instant Messaging Communications to Perform Author Identification for Cybercrime Investigations. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, *31 LNICST*, 99–110. https://doi.org/10.1007/978-3-642-11534-9_10.

Pennington, J., Socher, R., & Manning, C. (2014). {G}lo{V}e: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing ({EMNLP})* (pp. 1532–1543). Doha, Qatar: Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162.

Pstxy: Outlook .pst and .ost file reader for .Net. (n.d.). Retrieved February 14, 2021, from https://github.com/pantilesoft/pantilesoft.github.io.

Rafi, M. S. (2008). SMS Text Analysis: Language, Gender and Current Practices. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3827793.

Ron J, R. J. (n.d.). The Use Of Artificial Intelligence In Digital Forensics. Retrieved April 27, 2022, from https://www.exterro.com/blog/the-use-of-artificial-intelligence-in-digital-forensics.

Salah, R., & Gayar, N. El. (2019). Sentiment Analysis using Unlabeled Email data. In *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. EasyChair. Retrieved from https://ieeexplore.ieee.org/abstract/document/9004372.

Sentiment analysis - Wikipedia. (n.d.). Retrieved August 3, 2021, from https://en.wikipedia.org/wiki/Sentiment_analysis.

Shalini, K., Ravikurnar, A., Vineetha, R. C., Aravinda, R. D., Annd, K. M., & Soman, K. P. (2018). Sentiment Analysis of Indian Languages using Convolutional Neural Networks. *2018 International Conference on Computer Communication and Informatics, ICCCI 2018*. https://doi.org/10.1109/ICCCI.2018.8441371.

Shiha, M. O., & Ayvaz, S. (n.d.). The Effects of Emoji in Sentiment Analysis. https://doi.org/10.17706/ijcee.2017.9.1.360-369.

Sivasangari, V., Mohan, A. K., Suthendran, K., & Sethumadhavan, M. (2018). Isolating Rumors Using Sentiment Analysis. *Journal of Cyber Security and Mobility*, *7*(1), 181–200–181–200. https://doi.org/10.13052/2245-1439.7113.

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., … Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, *117*(30), 17680–17687. https://doi.org/10.1073/PNAS.1920484117.

Studiawan, H., Sohel, F., & Payne, C. (2020). Sentiment Analysis in a Forensic Timeline with Deep Learning. *IEEE Access*, *8*, 60664–60675. https://doi.org/10.1109/ACCESS.2020.2983435.

Technologies, Z. (n.d.-a). EDRM and ZL Launch New Enron Email Data Set – ZL Tech. Retrieved October 25, 2021, from https://www.zlti.com/press-releases/edrm-and-zl-launch-new-enron-email-data-set/.

Technologies, Z. (n.d.-b). Hillary Clinton’s Emails | Kaggle. Retrieved October 25, 2021, from https://www.kaggle.com/kaggle/hillary-clinton-emails?select=Emails.csv.

TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation. (n.d.). Retrieved October 10, 2021, from https://textblob.readthedocs.io/en/dev/.

Twitter US Airline Sentiment. (2019). Retrieved February 14, 2021, from https://www.kaggle.com/crowdflower/twitter-airline-sentiment/activity.

UCI Machine Learning Repository: SMS Spam Collection Data Set. (n.d.). Retrieved October 25, 2021, from https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection#.

Uses of AI in ediscovery | EDRM. (n.d.). Retrieved October 24, 2021, from https://edrm.net/wiki/3-uses-of-ai-in-ediscovery/.

Wang, Y., Kim, K., Lee, B., & Youn, H. Y. (2018). Word clustering based on POS feature for efficient twitter sentiment analysis. *Human-Centric Computing and Information Sciences 2018 8:1*, *8*(1), 1–25. https://doi.org/10.1186/S13673-018-0140-Y.

WordNet | A Lexical Database for English. (2010). Retrieved October 10, 2021, from https://wordnet.princeton.edu/.

Zhang, K., Cheng, Y., Xie, Y., Honbo, D., Agrawal, A., Palsetia, D., … Choudhary, A. (2011). SES: Sentiment elicitation system for social media data. In *Proceedings - IEEE International Conference on Data Mining, ICDM* (pp. 129–136). https://doi.org/10.1109/ICDMW.2011.153.

Zhang, L., Hua, K., Wang, H., Qian, G., & Zheng, L. (2014). Sentiment Analysis on Reviews of Mobile Users. *Procedia Computer Science*, *34*, 458–465. https://doi.org/10.1016/J.PROCS.2014.07.013.