

A SURVEY ON TEXT CLASSIFICATION USING DIFFERENT MACHINE LEARNING APPROACHES

Abstract

Since machine learning-based text categorization has so many practical applications, such as spam detection, hate speech identification, review summarization, rating analysis, sentiment modelling, and topic modelling, it is one of the most actively researched fields. Many different datasets, training methods, performance metrics, and comparison methods are used in popular machine learning studies. This paper's systematic review followed the guidelines laid out in the Preferred Reporting Items for Systematic Reviews (PRISMA) statement. Datasets, ML models, best accuracy, performance evaluation metrics, MT model comparisons, MT model splitting techniques, and MT model comparisons are all explored and analysed in the literature. We also point out omissions and make recommendations for further research in the field. While the reviewed works do a good job at classifying texts, there is room for improvement. We want for this literature review to serve as a resource for scholars and professionals in the field of text categorization.

Keywords: Text Classification, Machine Learning, Dataset, Metrics.

Authors

B. VeeraSekharreddy

Department of Information Technology
MLR Institute of Technology
Hyderabad, Telangana, India
Bhargavisekhar68@gmail.com

Venkata Nagaraju Thatha

Department of Information Technology
MLR Institute of Technology
Hyderabad, Telangana, India
Nagaraju.thatha@gmail.com

G. Uday Kiran

Department of Computer Science and
Engineering (AI&ML)
B V Raju Institute of Technology
Hyderabad, Telangana, India
udaykiran.goru@bvrit.ac.in

V. Srilakshmi

Department of Computer Science and
Engineering (AI&ML)
B V Raju Institute of Technology
Hyderabad, Telangana, India
Srilakshmi.v@bvrit.ac.in

Srinuvasarao Sanapala

Department of Computer Science and
Engineering
B V Raju Institute of Technology
Hyderabad, Telangana, India
ssrinuvas@gmail.com

I. INTRODUCTION

The explosion of digital textual data in recent years has given rise to novel insights and consequently new research fields. The stock market [1, 2]. Electronic news reports on the World Wide Web are already commonplace and have shown to be a vital data source [11]. However, the amount of news is tremendous, and it is not clear how to use it effectively for niche studies. Because of this, a domain-specific news monitoring system necessitates a framework or architecture, as well as a classification method for automatically categorizing relevant internet material into separate subject categories [5]. Monitoring the news is a form of quality control that checks and double-checks every news item that gets created, used, and stored. These techniques include processes for evaluating news to ensure its security, reliability, and credibility. There will always be a requirement for an approach that can sift relevant data from a large collection of textual materials covering various topics. Unstructured or semi-structured text constitutes the vast bulk of digital data sets [12]. Therefore, organizing this textual data became necessary to make it relevant for decision making [13, 14].

The sheer volume of information, however, makes manual processing unfeasible. This difficulty has driven advancements in text classification. It is the process of classifying texts into predetermined groups (called labels) based on their meaning and context. Most classification problems were previously done manually, which was time-consuming and costly at large scale. One way to look at classification is as the process of formulating guidelines for how to label materials based on their similarities. Some identifying information for a class is included in these rules. Although following handwritten rules is effective, it takes a lot of time and effort to create and update them over time. Rules can be framed by a technical expert in the form of regular expressions to improve the classifier's performance. There are a number of methods offered in the existing literature [15, 16] for automatically categorizing text texts using machine learning. With this method, the training data is used to discover automatically the rules or criteria for choosing a classifier. Numerous training records and knowledgeable labelers are needed for each category.

The labelling procedure consists of categorizing each document. Labelling was far simpler than developing rules from scratch. In addition, there are several supervised and semi-supervised learning methods that can lessen the need for human labelers [17, 18]. Automatic labelling can be used to accomplish this. Rule-based methods, data-driven methods, and hybrid methods are the three main categories of automated text classification techniques. Rule-based methods organize texts according to predetermined rules. Apple, grapes, and oranges are just a few examples of fruits that might get the "fruit" designation. These methods are complicated to maintain and call for in-depth familiarity with the subject matter. In contrast, data-driven approaches can be trained to create accurate classifications from existing data. The natural relationships between text and their labels can be learned by a machine learning algorithm utilizing labelled examples as training data. It is more adaptable and can be used for a variety of purposes; it is also capable of uncovering previously unseen patterns in the data. Hybrid methods combine the predictive power of both rule-based and machine learning (data-driven) techniques. Models of machine learning have garnered a lot of attention in recent decades [19, 20].

Extraction of features from text documents is the initial stage in most traditional machine learning models, with the features then being supplied to a classifier for prediction in a second step. BOW (bag-of-words) and TFIDF (term frequency-inverse document frequency) are two common feature representation approaches. Naive Bayes, k-nearest neighbors, support vector machines, decision trees, random forests, etc. are all examples of popular classifiers. The subsequent sections provide in-depth discussions of these paradigms. To enhance language modelling for broader context, researchers have applied deep learning models to a wide range of NLP applications [21–23]. End-to-end learning is being used by these models to attempt feature representation learning and categorization. They are much more adaptable from one project to the next and have the ability to reveal hidden patterns in the data. In recent years, these models have considerably shifted the paradigm for the numerous text classification tasks.

II. LITERATURE REVIEW

The task of text classification is defined as "the task of learning to distinguish among a set of predefined classes using features extracted from a collection of text documents" [29]. How accurately the training texts are split into classes and the classification granularity affect the classifier's accuracy [30, 31]. In text classification, we are provided with a set of labels or classes and asked to determine to which of those classes or labels a given textual material most closely conforms. A category or label is typically a broad subject, as "sports" or "business." Differentiating between articles on more comparable topics, including networks and the internet of things, may be substantially more challenging. It would be easier to learn if some traits, which signify the potential overlap between classes, were eliminated. In order to improve classification accuracy, it is necessary to widen the distance between classes. The process of features weighting and selecting useful features can help with this. The impact of sentiment classification on the stock market [31] are all examples of real-world applications of text classification.

To categorize a corpus of text, let's say we have N documents, and our first step is to construct a classifier, T . D is a set of written documents that have all been labelled or categorized by a human. The second step is to use a set of D documents that correspond to each class/label as input to train a classifier. It is time to put classifier C through its paces by using it to sort through N documents. Assigning a label or category to each document in N is a job for C . Data pretreatment, transformation, and dimensionality reduction are additional crucial processes in the text classification process that accompany model training. The first step is to gather information from diverse textual sources. The text may be part of a larger domain that describes specific activities, procedures, or facts. In order to provide an accurate text representation for the learning model, these text documents will need to undergo preprocessing. This is accomplished in two stages: first, features are retrieved from the processed text using any feature extraction algorithm, and then, features are reduced using feature selection techniques. The dimensions of data needed by the learning approach are reduced as a result of this simplification of the problem space. Following these steps, a set of learning algorithms is selected for use in supervised data training to produce a classifier with the highest likelihood of correctly identifying instances of a target class. Training data refers to the textual information needed to train a classifier. The bulk of the data is used to train the model, while a subset is used to evaluate the accuracy of the classifier (the "testing data"). In a similar vein, the model is educated to identify each target class based on the information included in the source texts. Classification methods are developed and tested by running them

on test data to determine the target class of input text; the result is then expressed as weights or probabilities. The accuracy of the text classifier is then assessed using evaluation methods. Here is the cutting edge research on text categorization.

An exhaustive survey of the field is available from Fabrizio Sebastiani [37]. The essay begins with a brief background on text classification and its importance in information retrieval and text analysis before moving on to other topics. Then, to improve classification accuracy and efficiency, Sebastiani explores various ways for encoding text documents, feature selection, and dimensionality reduction. Naive Bayes, Support Vector Machines (SVMs), and k-Nearest Neighbours (k-NN) are some of the supervised machine learning algorithms investigated while applying them to the problem of text classification. Challenges associated with large-scale and cross-lingual categorization are examined, and evaluation metrics to gauge system performance are introduced. Additionally, semi-supervised and active learning strategies are presented as ways to make use of unlabeled data in the article's examples.

The authors provide a thorough analysis of the many approaches taken in this area to help scholars and practitioners have a firm grasp on the subject as a whole. The review kicks out with a brief introduction to text classification, stressing its importance in IR and TM contexts. After that, they dig into methods of text preparation and feature extraction. Tokenization, stop-word elimination, and the popular TF-IDF (Term Frequency-Inverse Document Frequency) representation are all covered. These preprocessing methods are designed to boost the efficacy of future classification algorithms by improving the quality of features derived from textual input.

In [38], Kowaski investigates a wide variety of supervised text classification techniques. Naive Bayes, Support Vector Machines, k-Nearest Neighbours, and Decision Trees are some of the classic methods discussed here. Furthermore, the authors explore state-of-the-art deep learning techniques, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which have seen explosive growth in recent years due to their capacity to efficiently process sequential input, such as texts. Each algorithm's benefits and limitations are highlighted in the comparison, allowing readers to make more informed decisions when choosing methods for text classification.

The survey pays special emphasis to ensemble approaches, which use the combined power of numerous classifiers to boost performance. The authors elaborate on the usefulness of ensemble methods like Random Forest and AdaBoost for improving classification precision. In addition, the study focuses on the assessment measures used to evaluate text categorization models, illuminating the value of these metrics in gauging performance. The survey concludes with a look into domain-specific text classification, including the difficulties and potential benefits of adapting general text classification algorithms to specific fields like medicine, law, and social media. The authors wrap up by describing the current difficulties in text categorization, such as dealing with imbalanced datasets and multilingual texts, and suggesting avenues for future study to progress this important topic.

Rogers [39] discusses the importance and difficulties of real-time text classification on social media. They explore the repercussions of misclassification or delayed responses and stress the significance of quick processing of user-generated input. The methods used for social media sentiment analysis and topic recognition are the primary subject of this

systematic review of real-time text categorization methods. The authors provide an in-depth examination of the many machine learning algorithms, NLP strategies, and deep learning approaches that have been applied to the problem of real-time text classification. They shed light on the advantages and disadvantages of various methods, as well as their fitness for dealing with the peculiarities of social media data. The research also analyses the metrics and methods used to gauge the efficacy of real-time text classification systems. This critical analysis is useful for evaluating the claims made and contrasting the outcomes of different methods [40].

In this research, we explore how to apply deep reinforcement learning to the task of text summarization. In the field of machine learning known as reinforcement learning, agents are taught to act in response to information gleaned from their surroundings. The authors elaborate on how this strategy, by rewarding the model for creating relevant and brief content, can be used to generate high-quality and useful summaries. The research goes beyond deep reinforcement learning to investigate alternative cutting-edge approaches used in autonomous text summarization. It explores how NLP, deep neural networks, and attention mechanisms might be used to better collect contextual information and boost summarization performance. [41].

Automatic text summarization systems are evaluated based on a variety of performance criteria, and the authors of this article provide a critical study of these metrics and approaches. By comparing the proposed methodologies to other state-of-the-art approaches, this analysis clarifies the efficacy of the proposed approaches. This research sheds light on the use of deep reinforcement learning and other cutting-edge methods for automatically summarizing material. To better summarize massive amounts of textual data, it provides a comprehensive reference for researchers and practitioners interested in utilizing machine learning and natural language processing. Thangaraj et al. [42] looked at text categorization strategies in AI studies from 2010–2017 and categorized them based on their underlying algorithms. The report highlighted the benefits, drawbacks, and emerging trends in text categorization, and the findings were displayed in a tree structure to illustrate the connection between learning procedures and algorithms.

By identifying and analyzing central and cutting-edge works and goals in this field, Mironczuk et al. [43] provided an overview of the current status of text classification. The paper discussed six pillars of text classification and conducted a qualitative and quantitative analysis of the related publications. Text feature extraction, dimensionality reduction, existing algorithms, and assessment methods were briefly discussed in Kowsari et al. [44]. Both the strengths and weaknesses of each method, as well as their potential uses in solving real-world situations, were covered. Models for classifying texts were investigated by Wu et al. [18], who looked at CNNs, RNNs, Attention Mechanisms, and more. The limitations of prior approaches to text classification are outlined, and a deep-learning-based approach is introduced.

The technical contributions, similarities, and strengths of more than 150 deep-learning-based text categorization models were reviewed in detail by Minaee et al. [42]. The publication also explores potential avenues for further study and includes a description of more than 40 widely used text classification datasets. Data augmentation techniques for textual categorization were surveyed by Bayer et al. [28], who used a taxonomy to classify more than a hundred techniques into 12 broad categories. The report cites recent literature,

focuses on effective approaches, and looks ahead to potential areas of study. From 1961 through 2021, including models from classical to deep learning, Li et al. [43] discussed the state-of-the-art methods for text classification. A text classification taxonomy was created, and the research compared and contrasted various approaches, detailing their respective merits and shortcomings. Major ramifications, future research goals, and challenges in the field were also outlined.

III. RESEARCH GAPS

Several writers have tried, unsuccessfully, for years to use NLP techniques to extract meaning or identify semantic links between words in unstructured data [24,28]. However, when trying to construct a reliable categorization system, these approaches fall short. It permits the development of semi-supervised machine learning models, which require the manual labelling of some portions of training data while allowing machine learning algorithms to train on the remaining data [35]. Unseen writings are highly domain-specific, therefore these representations and universally available representations cannot be used to them [18]. Potential methods for expanding the vocabulary of existing representations for particular domains can be designed. However, these models are not equipped to deal with symbols directly, despite the fact that deep learning algorithms have been shown to be effective in decision-making for NLP-based applications. Training these kinds of algorithms also requires a lot of computing power. It provides room for the development of deep neural network-based architecture that can be fed lexical, linguistic, and subject-specific word information. There are some blank spots in our analysis. Some of them are

- Numerous academic investigations have highlighted individual-generated data sets. There are, nevertheless, great deals of datasets available. The data structures and formats of each of these collections are unique. Therefore, a multi-model categorization system needs to be created to deal with this problem.
- The best classifiers from the most popular datasets should be federated and stored in a publicly accessible database. Future researchers will be greatly aided by this in their efforts to produce high-quality results rapidly. This is a necessary step since it will allow researchers to check their findings against those of the larger population.
- Classification results can be enhanced with just a few inputs utilising active learning. Active learning is not used in the vast majority of the studies. This allows for the incorporation of active learning into the classification in subsequent works of research.
- The visualisation of text is still in need of refinement. Labelling or separating the traits should be the focus of several studies. A pronoun may be used to stand in for the noun it refers to in the preceding phrase. Therefore, it is important to have a proper labelling system.
- The majority of studies considered for this survey centred on a frequency-based ranking of characteristics. The most weight is given to the most common characteristics. However, this may be an exception rather than the rule. Experts in the field can help solve this problem by assigning a priority ranking to the least-used words.

IV. CONCLUSION

We have presented a comprehensive analysis of the text classification procedure in this work. In this study, we discussed several algorithms and strategies for doing categorization problems. It has outlined methods for gathering information from various online resources. To represent the documents, we first used some fundamental methods, then looked at what's new in document representation for NLP and machine learning, and so on. Data reduction from a higher to a lower dimensional space was achieved by feature selection and feature extraction techniques, allowing for the development of efficient and effective classifiers. The effectiveness of various techniques for training machine learning text classifiers varies widely across different domains. Based on their experience, the authors have found that none of these algorithms works perfectly for every problem scenario and dimension of data.

REFERENCES

- [1] Kapociute-Dzikiene, J.; Venckauskas, A.; Damasevicius, R. A comparison of authorship attribution approaches applied on the Lithuanian language. In Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, FedCSIS 2017, Prague, Czech Republic, 3–6 September 2017; pp. 347–351.
- [2] Mathews, A.; Sejal, N.; Venugopal, K.R. Text Based and Image Based Recommender Systems: Fundamental Concepts, Comprehensive Review and Future Directions. *Int. J. Eng. Trends Technol.* 2022, 70, 124–143.
- [3] Ji, Z.; Pi, H.; Wei, W.; Xiong, B.; Wozniak, M.; Damasevicius, R. Recommendation Based on Review Texts and Social Communities: A Hybrid Model. *IEEE Access* 2019, 7, 40416–40427.
- [4] Sun, G.; Wang, Z.; Zhao, J. Automatic text summarization using deep reinforcement learning and beyond. *Inf. Technol. Control* 2021, 50, 458–469.
- [5] Jiang, M.; Zou, Y.; Xu, J.; Zhang, M. GATSum: Graph-Based Topic-Aware Abstract Text Summarization. *Inf. Technol. Control* 2022, 51, 345–355.
- [6] Shrivvas, A.K.; Dewangan, A.K.; Ghosh, S.M.; Singh, D. Development of proposed ensemble model for spam e-mail classification. *Inf. Technol. Control* 2021, 50, 411–423.
- [7] Salloum, S.; Gaber, T.; Vadera, S.; Shaalan, K. A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. *IEEE Access* 2022, 10, 65703–65727.
- [8] Kapociute-Dzikiene, J.; Balodis, K.; Skadiene, R. Intent detection problem solving via automatic DNN hyperparameter optimization. *Appl. Sci.* 2020, 10, 7426.
- [9] Iqbal, W.; Malik, W.I.; Bukhari, F.; Almustafa, K.M.; Nawaz, Z. Big data full-text search index minimization using text summarization. *Inf. Technol. Control* 2021, 50, 375–389.
- [10] Dogra, V.; Verma, S.; Kavita; Chatterjee, P.; Shafi, J.; Choi, J.; Ijaz, M.F. A Complete Process of Text Classification System Using State-of-the-Art NLP Models. *Comput. Intell. Neurosci.* 2022, 2022, 1883698.
- [11] Ashokkumar, P.; Arunkumar, N.; Don, S. Intelligent optimal route recommendation among heterogeneous objects with keywords. *Comput. Electr. Eng.* 2018, 68, 526–535.
- [12] Haque, R.; Islam, N.; Tasneem, M.; Das, A.K. Multi-class sentiment classification on Bengali social media comments using machine learning. *Int. J. Cogn. Comput. Eng.* 2023, 4, 21–35.
- [13] Gupta, A.; Dengre, V.; Kheruwala, H.A.; Shah, M. Comprehensive review of text-mining applications in finance. *Financ. Innov.* 2020, 6, 39.
- [14] Li, Q.; Li, S.; Zhang, S.; Hu, J.; Hu, J. A review of text corpus-based tourism big data mining. *Appl. Sci.* 2019, 9, 3300.
- [15] Omoregbe, N.A.I.; Ndaman, I.O.; Misra, S.; Abayomi-Alli, O.O.; Damaševičius, R. Text messaging-based medical diagnosis using natural language processing and fuzzy logic. *J. Healthc. Eng.* 2020, 2020, 8839524.
- [16] Tesfagergish, S.G.; Damaševičius, R.; Kapociute-Dzikiene, J. Deep Fake Recognition in Tweets Using Text Augmentation, Word Embeddings and Deep Learning; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2021; Volume 12954, pp. 523–538.

- [17] Thangaraj, M.; Sivakami, M. Text Classification Techniques: A Literature Review. *Interdiscip. J. Inf. Knowl. Manag.* 2018,13, 117–135.
- [18] Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep Learning–based Text Classification. *ACM Comput. Surv.* 2021, 54, 1–40.
- [19] Bayer, M.; Kaufhold, M.A.; Reuter, C. A Survey on Data Augmentation for Text Classification. *ACM Comput. Surv.* 2022,55, 3544558.
- [20] Li, Q.; Peng, H.; Li, J.; Xia, C.; Yang, R.; Sun, L.; Yu, P.S.; He, L. A Survey on Text Classification: From Traditional to Deep Learning. *ACM Trans. Intell. Syst. Technol.* 2022, 13, 1–41.
- [21] Wu, H.; Liu, Y.; Wang, J. Review of text classification methods on deep learning. *Comput. Mater. Contin.* 2020, 63, 1309–1321.
- [22] A. Atkins, M. Niranjani, and E. Gerding, “Financial news predicts stock market volatility better than close price,” *Me Journal of Finance and Data Science*, vol. 4, no. 2, pp. 120–137, 2018.
- [23] J. Robinson, A. Glean, and W. Moore, “How does news impact on the stock prices of green firms in emerging markets?” *Research in International Business and Finance*, vol. 45, pp. 446–453, 2018.
- [24] G. L’offler, L. Norden, and A. Rieber, *Salient News and the Stock Market Impact of Tone in Rating Reports*, Ssrn, Rochester, NY, USA, pp. 1–41, 2016.
- [25] M. N. Elagamy, C. Stanier, and B. Sharp, “Stock market random forest-text mining system mining critical indicators of stock market movements,” in *Proceedings of the 2018 Second International Conference on Natural Language and Speech Processing (ICNLSP)*, pp. 1–8, IEEE, Algiers Algeria, June 2018.
- [26] Y. Zhang, Y. Dang, H. Chen, M. /urmond, and C. Larson, “Automatic online news monitoring and classification for syndromic surveillance,” *Decision Support Systems*, vol. 47, no. 4, pp. 508–517, 2009.
- [27] J. C. West, D. E. Clarke, F. F. Duffy et al., “Availability of mental health services prior to health care reform insurance expansions,” *Psychiatric Services*, vol. 67, no. 9, pp. 983–989, 2016.
- [28] M. Ali, S. Khalid, M. I. Rana, and F. Azhar, “A probabilistic framework for short text classification,” in *Proceedings of the IEEE Eighth AnnuComputCommun Work Conf CCWC 2018*, pp. 742–747, Las Vegas, NV, USA, February 2018.
- [29] C. Romero and S. Ventura, “Educational data mining: a survey from 1995 to 2005,” *Expert Systems with Applications*, vol. 33, no. 1, pp. 135–146, 2007.
- [30] Q. Ye, Z. Zhang, and R. Law, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 6527–6535, 2009.
- [31] D. /orleuchter and D. Van Den Poel, “Predicting e-commerce company success by mining the text of its publiclyaccessible website,” *Expert Systems with Applications*, vol. 39, no. 17, Article ID 13026, 2012.
- [32] G. Yang, M. A. Jan, A. U. Rehman, M. Babar, M. M. Aimal, and S. Verma, “Interoperability and data storage in Internet of multimedia things: investigating current trends, research challenges and future directions,” *IEEE Access*, vol. 8, Article ID 124382, 2020.
- [33] A. Gandomi and M. Haider, “Beyond the hype: big data concepts, methods, and analytics,” *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [34] Y. Sun, A. K. C. Wong, and M. S. Kamel, “Classification of imbalanced data: a review,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 4, pp. 687–719, Lyon, France, 2009.
- [35] S. Manne and S. S. Fatima, “A novel approach for text categorization of unorganized data based with information extraction,” *International Journal of Computational Science and Engineering*, vol. 3, pp. 2846–2854, 2011.
- [36] B. S. Harish, D. S. Guru, and S. Manjunath, “Representation and classification of text documents: a brief review,” *IJCA, Spec Issue Recent Trends Image Process Pattern Recognit*, pp. 110–119, 2010.
- [37] Y. Liang, Y. Liu, C. K. Kwong, and W. B. Lee, “Learning the “Whys”: discovering design rationale using text mining – an algorithm perspective,” *Computer-Aided Design*, vol. 44, no. 10, pp. 916–930, 2012.
- [38] B. Liu, X. Li, W. S. Lee, and P. S. Yu, “Text classification by labeling words,” *Artificial Intelligence*, vol. 34, pp. 425–430, 2004.
- [39] D. Y. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Scholkopf, “Learning with local and global consistency,” *Advances in Neural Information Processing Systems*, vol. 16, pp. 321–328, 2004.
- [40] H. U. Rehman, R. Rafique, and M. C. M. Nasir, “Forecasting CO2 emissions from energy, manufacturing and transport sectors in Pakistan: statistical vs. Machine learning methods,” *Mach Learn Methods*, vol. 28, p. 21, 2017.
- [41] N. Kundu, G. Rani, V. S. Dhaka et al., “IoT and interpretable machine learning based framework for disease prediction in pearl millet,” *Sensors*, vol. 21, no. 16, p. 5386, 2021.
- [42] Y. Bengio, R. Ducharme, and P. Vincent, “A neural probabilistic language model,” *Advances in Neural Information Processing Systems*, vol. 3, pp. 1–11, 2001.

- [43] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [44] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015