International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-32, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Enhancing Customer Service in Banking with AI: Intent Classification Using Distilbert

Saurabh Kumar[1], Suman Deep[2], Pourush Kalra[3]
[1] Sr Data Science Manager, CA, USA
[2] Technical Architect, CA, USA
[3] Business Operations Associate, CA, USA

ABSTRACT: With the increasing demand for efficient and responsive customer service in the banking sector, artificial intelligence offers a promising solution. This paper presents a comparative analysis of artificial intelligence methodologies applied to intent classification within the banking sector customer service domain. Utilizing a comprehensive dataset of banking service inquiries, we evaluate several machine learning approaches, including Naive Bayes, Logistic Regression, Support Vector Machine with Linear Kernel, Random Forest, XGBoost, and the transformer-based DistilBERT model. The models are assessed based on their accuracy, precision, recall, and F1 score metrics. Our findings indicate that DistilBERT, with its distilled architecture, not only outstrips traditional models but also demonstrates exceptional performance with an accuracy and F1 score exceeding 92%. The paper delves into the advantages of employing such an efficient and powerful model in real-time customer service settings, suggesting that DistilBERT offers a substantial enhancement over conventional methods. By providing detailed insights into the model's capabilities, we underscore the transformative impact of employing advanced AI in the financial industry to elevate customer service standards, streamline operational efficiency, and harness the power of state-of-the-art technology for improved client interactions. The results showcased in this study are indicative of the strides being made in AI applications for financial services and set a benchmark for future exploratory and practical endeavors in the field.

KEYWORDS: Banking Sector, DistilBERT, Intent Classification, Natural Language Processing (NLP), Transformer Models.

1. INTRODUCTION

In the dynamic realm of customer service, Artificial Intelligence (AI) has become a cornerstone, significantly enhancing the efficiency and effectiveness of business-client interactions. Central to AI-driven customer service is the task of intent classification—a critical component for discerning and addressing customer inquiries with precision. Traditional NLP methodologies have primarily utilized machine learning algorithms such as Naive Bayes, SVM, and logistic regression to tackle this challenge. However, with the introduction of deep learning, specifically the advent of transformer models like BERT and its variants, the domain has witnessed a paradigm shift [1][2].

This research places its emphasis on DistilBERT, a more compact derivative of BERT that conserves the majority of its original model's functionality while optimizing for speed and minimal resource utilization [3]. The adoption of DistilBERT is particularly salient in customer service operations, where the constraints of resources and the necessity for real-time processing intersect. Through the application of DistilBERT to the Banking77 dataset, a stringent benchmark for nuanced intent detection in the banking field, our analysis shows that DistilBERT notably exceeds the capabilities of classical ML models [6].

The incorporation of DistilBERT offers a tangible solution for the practical application within customer service scenarios. It not only fosters increased accuracy and expediency in intent classification but also improves customer interactions and satisfaction by enabling quicker and more accurate responses to inquiries.

In addition to these practical applications, our study evaluates DistilBERT against traditional ML techniques that employ the term frequency-inverse document frequency (TF-IDF) method to train models—a long-standing practice in NLP[33][34]. This comparison is crucial as it allows us to forego the creation of linguistically-detailed resources that would require expert management, substituting it with an approximation via the TF-IDF scoring for each term. While this may lead to a slight decrease in precision, it is counterbalanced by a gain in recall, thus expanding the pool of relevant results that can be retrieved.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-32, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Furthermore, we explore how the latest enhancements in training approaches, like transfer learning and few-shot learning, can refine DistilBERT's performance within specialized sectors [5][7][8]. Our ambition is to deliver a scalable, efficient AI tool that not only augments the quality of customer service but also lays the groundwork for broader AI adoption across various industries.

2. LITERATURE REVIEW AND CHALLENGES

The task of intent classification has been a crucial focus in the field of natural language understanding (NLU), with significant advancements in recent years. Traditional machine learning approaches, such as Naive Bayes, XGBoost, Random Forest, Support Vector Machines (SVM), and Logistic Regression, have been widely employed for intent classification [9][10][11]. However, these methods often struggle to capture the nuances and complexities of natural language, leading to suboptimal performance.

The advent of deep learning and transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers) [1], has revolutionized the field of NLU. BERT's self-attention mechanism and bidirectional training approach have enabled it to capture contextual information more effectively, leading to state-of-the-art performance on various NLP tasks, including intent classification [1][2]. However, BERT's large model size and computational requirements have posed challenges for real-world deployment, particularly in resource-constrained environments.

To address these limitations, researchers have explored distilled versions of BERT, such as DistilBERT [3], which aims to retain the performance of the original model while reducing its size and computational complexity. DistilBERT leverages knowledge distillation techniques to transfer the knowledge from the larger BERT model to a smaller student model, resulting in a more efficient and lightweight architecture [3].

Several studies have investigated the effectiveness of DistilBERT for intent classification tasks. Casanueva et al. [31] proposed a dual sentence encoder approach using DistilBERT for efficient intent detection. Larson et al. [4] introduced an evaluation dataset for intent classification and out-of-scope prediction, demonstrating the potential of DistilBERT in handling diverse intents. Additionally, researchers have explored the application of DistilBERT in various domains, such as legal contract review, conversational recommendation [30], and few-shot learning [7][8].

The banking domain has emerged as a crucial area for intent classification, as accurate intent detection is essential for delivering exceptional customer service. The banking77 dataset from Hugging Face [6] provides a unique opportunity to evaluate the performance of DistilBERT in a real-world setting, where the model's ability to handle fine-grained intents and domain-specific terminology is crucial.

This literature review highlights the importance of intent classification, the limitations of traditional machine learning approaches, and the potential of transformer-based models, particularly DistilBERT, in addressing these challenges. By leveraging the banking77 dataset and comparing DistilBERT's performance with traditional methods, this research aims to contribute to the ongoing efforts in enhancing customer service through AI-driven intent classification.

2.1 Challenges

Despite advances in AI for intent classification, significant challenges remain:

- **Computational Complexity:** Deploying models like DistilBERT in resource-constrained environments is challenging due to their substantial computational demands, hindering real-time applications [3].
- **Data Privacy and Security:** Training NLP models in the banking sector must comply with strict data protection regulations, complicating the use of sensitive customer data.
- **Integration with Existing Systems:** Integrating modern NLP technologies into legacy banking systems requires significant technical adaptation and investment [4].
- **Handling Imbalanced Data:** Imbalanced datasets can skew model accuracy, necessitating advanced techniques to ensure equitable intent recognition across classes [32].

3. DATASET DESCRIPTION

The dataset utilized for this study is the Banking77 dataset.The Banking77 dataset from Hugging Face provides an advanced benchmark for intent detection in NLP research. It consists of 13,083 banking-related queries categorized into 77 distinct intents, reflecting the intricate nature of customer service interactions [6]. The dataset is designed to mirror the

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-32, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Complexity and variability of real-world banking queries, thus providing a robust platform for training and evaluating the DistilBERT model's classification capabilities.

4. METHODOLOGY

We will outline traditional NLP methods for comparison with DistilBERT, and then delve into the specifics of the BERT and DistilBERT models.

4.1 Traditional Classification Model with TF-IDF

A foundational method for text classification in supervised learning NLP tasks is the utilization of the bag-of-words model, augmented by the Term Frequency-Inverse Document Frequency (TF-IDF) weighting.The methodology incorporates the computation of the TF-IDF to transform textual information into a feature vector that represents the importance of words within the documents [35].TF-IDF is a statistical measure used to evaluate the importance of a word to a document in a collection or corpus. The term frequency (TF) denotes the number of times a word appears in a document, normalized over the document length. The inverse document frequency (IDF) component diminishes the weight of terms that occur very frequently across the corpus and increases the weight of terms that occur rarely**.**

The term frequency (TF) is calculated as:

$$\mathrm{TF}(t_i, d) = \frac{\text{Number of times term } t_i \text{ appears in document } d}{\text{Total number of terms in document } d}$$

The inverse document frequency (IDF), which measures the informativeness of term

$$\mathrm{IDF}(t_i, D) = \log\left(\frac{N}{n_i}\right)$$

Where N is the total number of documents in the corpus D and n_i is the number of documents where the term t_i appears.

The TF-IDF weight is the product of these two figures:

$$\text{TF-IDF}(t_i, d, D) = \mathrm{TF}(t_i, d) \times \mathrm{IDF}(t_i, D)$$

This weight is high when t_i is frequent in a small number of documents, thus lending high discriminating power to the term.

Following the creation of the TF-IDF matrix [35], traditional machine learning classifiers, such as Naive Bayes, Support Vector Machines (SVM), or Logistic Regression, are trained using this matrix as input. The models are assessed using standard evaluation metrics to determine their predictive power on the Banking77 dataset.

4.2 Transformer-Based Models: BERT and DistilBERT

4.2.1 BERT

In advancing the frontiers of NLP, the BERT (Bidirectional Encoder Representations from Transformers) model has established itself as a paradigm-shifting methodology. The process comprises two distinct phases: pre-training and fine-tuning [1]. Initially, BERT undergoes pre-training on a large unlabeled text corpus, learning a general representation of language by predicting masked words and the relationships between sentences. Subsequently, the model is fine-tuned, utilizing the pre-trained parameters as a starting point and adjusting all parameters based on labeled data specific to the target task.

The architecture of BERT is rooted in a multi-layer bidirectional transformer encoder [1][2]. It is composed of a series of layers, each containing two distinct sub-layers: a multi-head self-attention mechanism and a position-wise fully connected feedforward network. The innovative design

Incorporates a residual connection around each sub-layer, followed by layer normalization [2]. This design enables the model to integrate information from both the preceding and following context dynamically across the input sequence.

The attention mechanism central to BERT is scaled dot-product attention, which operates on queries, keys, and values represented by matrices Q, K, and V, respectively. It computes the attention scores as:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

The model employs multiple such attention heads to capture various aspects of the input's semantic and syntactic features, which are then concatenated to form the final output [2].

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-32, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Pre-training tasks for BERT include a masked language model (LM) task, where random tokens are masked and predicted by the model, and a next sentence prediction task that teaches BERT to understand sentence relationships [1]

Fine-tuning adapts BERT's extensive pre-trained knowledge to a specific downstream task, leveraging the self-attention mechanism to model diverse tasks with minimal task-specific modifications. The inputs and outputs are tailored for each task, with all parameters adjusted to optimize performance [1].

4.2.2 DistilBERT

DistilBERT is a streamlined version of the BERT
architecture, designed to be more suitable for use in settings with limited computational resources. It maintains a structure similar to BERT but is notable for having 40% fewer trainable parameters, striking a balance between efficiency and performance, allowing it to retain about 97% of BERT's effectiveness in various natural language processing (NLP) tasks [15].

The design of DistilBERT includes six transformer blocks, 768 hidden units, and 12 self-attention heads, mirroring BERT's setup. This configuration helps it to perform nearly as well as BERT while being more compact, making it both scalable and capable of processing data in real time without significantly compromising results [15].

While there are other BERT variants like RoBERTa, TinyBERT, and ALBERT that focus on improving performance, DistilBERT's key advantage is its ability to provide nearly the same accuracy as BERT with a much
smaller model size. This makes it especially useful for real-time applications and on devices with limited computing capabilities [3][10][27]

Table 1: Comparison of BERT and DistilBERT

Parameters	BERT (base)	DistilBERT(base)
Trainable Parameters	110M	66M
No. of Layers	12	6
No. of hidden units	768	768
No. of self attention heads	12	12

The methodology for implementing DistilBERT follows the standard of fine-tuning for specific tasks. This involves adapting the pre-trained parameters to the targeted dataset, in this case, Banking77, to accurately classify customer service intents with the efficiency required in an operational setting [6].

4.3 Model Evaluation Metrics

1. Accuracy: Accuracy measures the proportion of total correct predictions (both true positives and true negatives) among the total number of cases examined.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

2. Precision (Positive Predictive Value) : Precision is the ratio of correctly predicted positive observations to the total predicted positives. It is a measure of a
3. Classifier's exactness. High precision relates to a low false positive rate.

$$\text{Precision} = \frac{TP}{TP + FP}$$

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-32, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

4. Recall (Sensitivity, True positive rate) : Recall is the ratio of correctly predicted positive observations to all observations in actual class. It is a measure of a classifier's completeness.

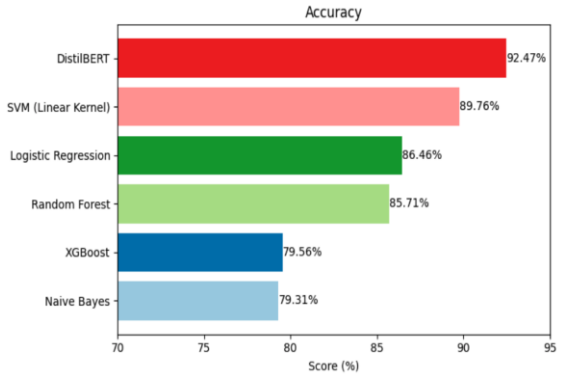
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

5. F1 Score: The F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is particularly useful when the class distribution is uneven.

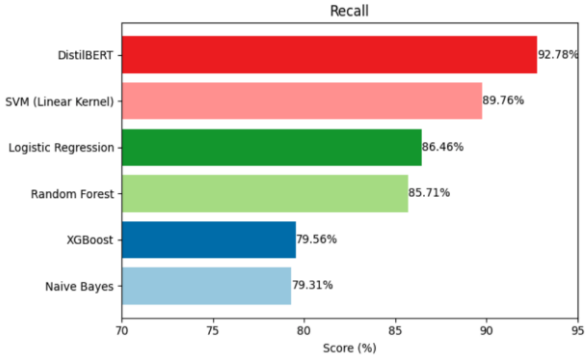
$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. RESULTS

The results of the study confirm the efficacy of DistilBERT in handling complex text classification tasks, outperforming traditional models such as SVM, Logistic Regression, Random Forest, Naive Bayes, and even XGBoost across various metrics including Accuracy, Precision, Recall, and F1 Score. Particularly noteworthy is the precision of DistilBERT, which reached a remarkable 92.48%, alongside its recall rate at 92.78% and an F1 score of 92.47%, solidifying its position as the leading model. These results suggest that when it comes to nuanced tasks like intent classification within customer service, DistilBERT provides a reliable and efficient solution. The study's focus on comparing DistilBERT's performance on the Banking77 dataset elucidates the model's potential to handle fine-grained intent detection in real-world customer service scenarios.

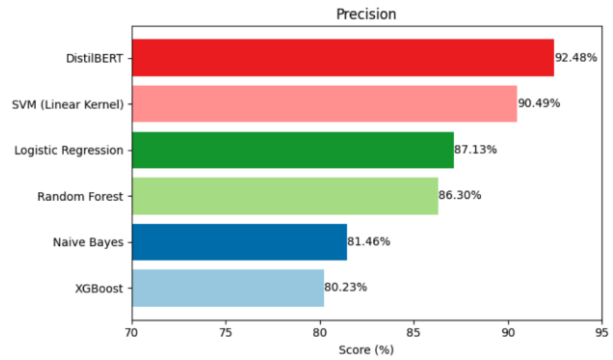


1a) Accuracy

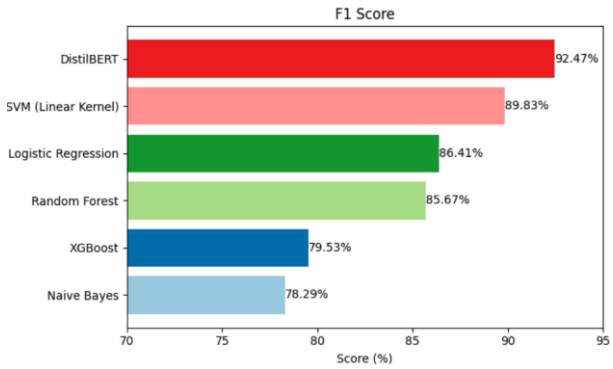


1b) Recall

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-32, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org



1c) Precision



1d) F1 Score

Fig.1 Comparison of model performance metrics

By optimizing for high accuracy and recall, DistilBERT ensures that customer intents are not only accurately identified but also that very few relevant intents are missed. The high precision indicates a low rate of false positives, which is crucial for maintaining customer satisfaction by avoiding misdirected or irrelevant responses. The F1 Score, which balances precision and recall, confirms that DistilBERT maintains this balance excellently, making it an ideal choice for deployment in customer service environments where both understanding the customer's intent and providing accurate responses quickly are essential. This study lays the groundwork for further research into the applicability of DistilBERT and similar models within various domains of the service industry, and their potential to revolutionize customer service through AI-driven solutions

6. CONCLUSION

Enhancing customer service within the banking sector using the DistilBERT model for intent classification is the focus of this research. The application of DistilBERT has proven to significantly enhance the understanding and processing of customer inquiries, thereby boosting the efficiency and accuracy of responses in real-time settings. This research demonstrates that DistilBERT not only excels traditional machine learning models such as Naive Bayes, SVM, Logistic Regression, Random Forest, and XGBoost with an impressive accuracy of 92.47% and an F1 score of 92.47%, but also maintains high precision and recall rates exceeding 92%.

In this study, a meticulous comparison of DistilBERT's performance against traditional classifiers underscores its superior capability to handle fine-grained intent detection, which is essential for effective customer service in the banking industry. The achievement of the research is quantified by the notable improvement in performance metrics, affirming that DistilBERT offers a substantial enhancement over conventional methods.

Future work may explore the integration of DistilBERT within more complex, real-time customer interaction systems across different sectors to further validate its effectiveness and adaptability. Enhancements in training approaches and optimization techniques could also amplify its performance, paving the way for broader AI adoption and operational excellence in customer service domains.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-32, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

REFERENCES

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. [CrossRef] [Google Scholar] [Publisher Link]
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.[CrossRef] [Google Scholar] [Publisher Link]
3. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.[CrossRef] [Google Scholar] [Publisher Link]
4. Larson, S., Mahendran, A., Peper, J. J., Clarke, C., Lee, A., Hill, P., ... & Mars, J. (2019). An evaluation dataset for intent classification and out-of-scope prediction. arXiv preprint arXiv:1909.02027.[CrossRef] [Google Scholar] [Publisher Link]
5. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research, 21(140), 1-67.[CrossRef] [Google Scholar] [Publisher Link]
6. Hugging Face. (2022). Banking77 [Data set]. Hugging Face Datasets.
7. Schick, T., & Schütze, H. (2020). Exploiting cloze questions for few shot text classification and natural language inference. arXiv preprint arXiv:2001.07676.[CrossRef] [Google Scholar] [Publisher Link]
8. Gao, T., Fisch, A., & Chen, D. (2020). Making pre-trained language models better few-shot learners. arXiv preprint arXiv:2012.15723.[CrossRef] [Google Scholar] [Publisher Link]
9. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.[CrossRef] [Google Scholar] [Publisher Link]
10. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.[CrossRef] [Google Scholar] [Publisher Link]
11. Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).[CrossRef] [Google Scholar] [Publisher Link]
12. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the association for computational linguistics, 5, 135-146. [CrossRef] [Google Scholar] [Publisher Link]
13. Peters, M. E., Neumann, M., Zettlemoyer, L., & Yih, W. T. (2018). Dissecting contextual word embeddings: Architecture and representation. arXiv preprint arXiv:1808.08949. [CrossRef] [Google Scholar] [Publisher Link]
14. Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.[CrossRef] [Google Scholar] [Publisher Link]
15. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. [CrossRef] [Google Scholar] [Publisher Link]
16. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32.[CrossRef] [Google Scholar] [Publisher Link]
17. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.[CrossRef] [Google Scholar] [Publisher Link]
18. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.[CrossRef] [Google Scholar] [Publisher Link]
19. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.[CrossRef] [Google Scholar] [Publisher Link]
20. Yin, W., Hay, J., & Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. arXiv preprint arXiv:1909.00161. [CrossRef] [Google Scholar] [Publisher Link]
21. González-Carvajal, S., & Garrido-Merchán, E. C. (2020). Comparing BERT against traditional machine learning text classification. arXiv preprint arXiv:2005.13012.[CrossRef] [Google Scholar] [Publisher Link]

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-32, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

22. Bowman, S. R., & Dahl, G. E. (2021). What will it take to fix benchmarking in natural language understanding?. arXiv preprint arXiv:2104.02145. [CrossRef] [Google Scholar] [Publisher Link]
23. Sun, S., Cheng, Y., Gan, Z., & Liu, J. (2019). Patient knowledge distillation for bert model compression. arXiv preprint arXiv:1908.09355. [CrossRef] [Google Scholar] [Publisher Link]
24. Phang, J., Févry, T., & Bowman, S. R. (2018). Sentence encoders on stilts: Supplementary training on intermediate labeled-data tasks. arXiv preprint arXiv:1811.01088. [CrossRef] [Google Scholar] [Publisher Link]
25. Tang, R., Lu, Y., Liu, L., Mou, L., Vechtomova, O., & Lin, J. (2019). Distilling task-specific knowledge from bert into simple neural networks. arXiv preprint arXiv:1903.12136. [CrossRef] [Google Scholar] [Publisher Link]
26. Vu, T., Wang, T., Munkhdalai, T., Sordoni, A., Trischler, A., Mattarella-Micke, A., ... & Iyyer, M. (2020). Exploring and predicting transferability across NLP tasks. arXiv preprint arXiv:2005.00770.[CrossRef] [Google Scholar] [Publisher Link]
27. Pruksachatkun, Y., Phang, J., Liu, H., Htut, P. M., Zhang, X., Pang, R. Y., ... & Bowman, S. R. (2020). Intermediate-task transfer learning with pretrained models for natural language understanding: When and why does it work?. arXiv preprint arXiv:2005.00628.[CrossRef] [Google Scholar] [Publisher Link]
28. Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., ... & Liu, Q. (2019). Tinybert: Distilling bert for natural language understanding. arXiv preprint arXiv:1909.10351. [CrossRef] [Google Scholar] [Publisher Link]
29. Hendrycks, D., Burns, C., Chen, A., & Ball, S. (2021). Cuad: An expert-annotated nlp dataset for legal contract review. arXiv preprint arXiv:2103.06268. [CrossRef] [Google Scholar] [Publisher Link]
30. Liu, Z., Wang, H., Niu, Z. Y., Wu, H., Che, W., & Liu, T. (2020). Towards conversational recommendation over multi-type dialogs. arXiv preprint arXiv:2005.03954. [CrossRef] [Google Scholar] [Publisher Link]
31. Casanueva, I., Temčinas, T., Gerz, D., Henderson, M., & Vulić, I. (2020). Efficient intent detection with dual sentence encoders. arXiv preprint arXiv:2003.04807.[CrossRef] [Google Scholar] [Publisher Link]
32. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284. [CrossRef] [Google Scholar] [Publisher Link]
33. Yun-tao, Z., Ling, G., & Yong-cheng, W. (2005). An improved TF-IDF approach for text classification. Journal of Zhejiang University-Science A, 6(1), 49-55. [CrossRef] [Google Scholar] [Publisher Link]
34. Trstenjak, B., Mikac, S., & Donko, D. (2014). KNN with TF-IDF based framework for text categorization. Procedia Engineering, 69, 1356-1364. [CrossRef] [Google Scholar] [Publisher Link]
35. Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. Journal of documentation, 60(5), 503-520. [CrossRef] [Google Scholar] [Publisher Link]

Cite this Article: Saurabh Kumar, Suman Deep, Pourush Kalra (2024). Enhancing Customer Service in Banking with AI: Intent Classification Using Distilbert. International Journal of Current Science Research and Review, 7(5), 2706-2713