Neural Text Classification for Digital Transformation in the Financial Regulatory Domain

Nelson Correa
Andinum, Inc.
West Palm Beach, FL, USA
ncorrea@ieee.org

Antonio Correa
Andinum, Inc.
Weston, FL, USA
antonio@andinum.com

***Abstract* — A core use case in artificial intelligence and natural language processing (NLP) is automatic *text classification* of documents, for the efficient, transparent and reliable handling of the billions of documents generated each year as part of business and government operation. Our application for document analysis uses deep learning for Neural Text Classification, with recurrent (*Bi-LSTM*) and transformer neural networks (*DistilBERT* and *FinBERT*). We compare the new models against traditional TF-IDF bag-of-words machine learning models, and evaluate text classification on a corpus of over 2,600,000 consumer financial complaints from the U.S. Consumer Financial Protection Bureau (CFPB), an agency of the U.S. Federal government created as a result of the 2008 financial crisis. Our analysis shows the superiority of the transformer models, with a classification accuracy of 88.05% on the task formulated.**

Keywords — Text Classification, Digital Transformation, Financial Regulation, Machine Learning, Deep Learning, Natural Language Processing

I. INTRODUCTION

Recent advances in artificial intelligence, machine learning and natural language processing (AI/ML/NLP) have enabled the development and deployment of business applications for tasks of sensory perception (vision, speech, sound, touch) and tasks that require a high level of cognitive ability (language understanding, visual scene analysis, reasoning and decision-making) for applications in the real and the virtual world [1].

Businesses and government are run on laws, rules and regulations, captured in millions of documents, that in turn generate billions of other documents every year, as a result of businesses and government operation. Text classification [2, 3] and other tasks in natural language processing find wide use in digital business transformation [4], for automation of document and other business processes.

In this paper we present consumer financial complaint classification on a corpus of over 2,600,000 complaints with "*product*" and "*issue*" labels, from the U.S. Consumer Financial Protection Bureau (CFPB) [5], an agency of the U.S. Federal government created as a result of the 2008 financial crisis (Dodd-Frank Act of 2010) [6].

In this work we use recent deep learning models for Neural Text Classification, including recurrent [7, 8], attention [9] and transformer neural networks [10]. One of our aims in the paper is to compare the performance of the new neural models on the CFPB dataset against classical information retrieval methods [11], with traditional TF-IDF weighted bag-of-words machine learning models. The characteristics and classification accuracy of the baseline and neural models developed are discussed, obtaining 88.05% accuracy on the classification task.

The contributions of this paper are:

1. Description of the CFPB dataset and recent neural models for text classification;

2. Evaluation of five machine learning models for text classification on the CFPB dataset;

3. A transformer large language model with an accuracy of 88.05%, that outperforms previous models and is suitable for automatic classification of CFPB complaints.

The paper is not only technical, but instead also presents the business problem, the technical approach and methodology common to the execution of modern machine learning and artificial intelligence projects, and the opportunities and key societal and ethical implications of development of these new technologies.

II. CFPB CONSUMER COMPLAINTS DATABASE

The CFPB consumer complaints database has been continuously collected since the end of 2011, with approximately 994,000 complaints received in 2021 alone, and over 2,600,000 complaints (2.6 GB data) total.

In this paper we use a previously-made available version of the CFPB dataset from Kaggle, including data up to 2016 [12]. This subset of the CFPB dataset contains 555,957 records, with 66,806 records with non-empty complaint document texts.

A. Data fields and consumer text narratives

The CFPB database is available in CSV and JSON data formats. We use the CSV format. Each record contains 18 fields, with key fields "*complaint_id*", "*date_received*", "*complaint_what_happend*", "*company*", "*state*" and "*zip_code*" and label fields "*product*", "sub-*product*", "*issue*" and "*sub-issue".* The consumer complaint narrative is "*complaint_what_happend*", and the data labels "*product*" and "*issue*", may serve as target classification labels (Fig. 1).

complaint_id	complaint_what_happened	product	issue
4971409	I have already filed a dispute on the incorrec...	Credit reporting, credit repair services, or o...	Problem with a credit reporting company's inve...
5037553	On XX/XX/21 at approximately XXXX am I was con...	Debt collection	Communication tactics
5020621	Hello, I have a loan with PNC bank, and it has...	Vehicle loan or lease	Managing the loan or lease
5014775	I had a car loan with XXXX XXXX. There was a C...	Credit reporting, credit repair services, or o...	Problem with a credit reporting company's inve...

Fig. 1. Sample CFPB database records

The consumer complaint narrative on the subset used here has 66,806 documents, 12,736,164 words, and a vocabulary of 49,451 unique words, with mean document length of 190.6 words (minimum of 2 and maximum of 1,284 words). The distribution histogram of complaint length in words is shown in Fig. 2.

978-1-6654-8854-9/22/$31.00 ©2022 IEEE

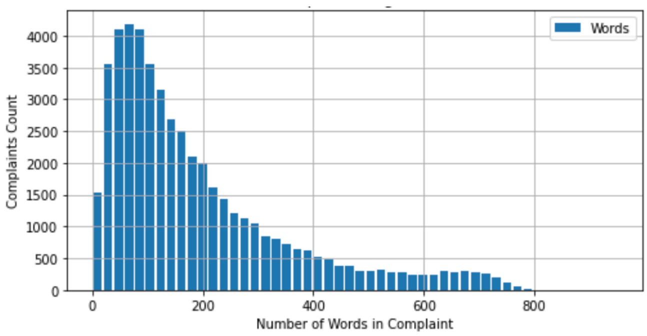

Fig. 2. CFPB Complaint length, in words

The following is a sample extract of a complaint record with a short narrative.

```
"I became a victim of identity theft
couple of years ago. After that incident I
have noticed many incorrect and unauthorized
items appeared in my report. So, I am
requesting that you delete all the following
ACCOUNTS AND INQUIRIES from my credit report
immediately."
```

It is important to note that consumer complaint narratives in the publicly released CFPB database are included only with consumer consent and after taking steps to remove sensitive information (names, addresses, account numbers and other personally identifiable information).

In this paper we focus on the fields "*complaint_id*", "*complaint_what_happend*" and "*product*", and comment briefly on the "*issue*" field.

B. Data labels: "Product" and "Issue"

Complaints in the CFPB database are classified in a two-level hierarchy of "*product*", "sub-*product*" and "*issue*", "*sub-issue*". We describe "*product*" and "*issue*" next.

The "*product*" label includes 13 consumer financial product and service categories (collectively "product"): Credit or consumer reporting; Debt collection; Credit card; Checking or savings account; Mortgage; Money transfer, money services, and virtual currencies; Vehicle loan or lease; Prepaid card; Student loan; Payday loan , title loan, or personal loan; and Credit repair.

For the purposes of this work, there are eleven "*Product*" categories of the original thirteen in the data subset used.

With this, the "*Product*" categories are:

1. Bank account or service
2. Consumer Loan
3. Credit card
4. Credit reporting
5. Debt collection
6. Mortgage
7. Student loan
8. Money transfers
9. - 11. Payday loan, Prepaid card, Other financial service

Complaints in the CFPB database are also classified into over 90 "*issue*" categories, including (top 5): "*Problem with a credit reporting company*", "*Incorrect information on your report*","*Improper use of your report*", "*Attempts to collect debt not owed*", and "*Managing an account*".

For data analysis, the "*issue*" categories may be reduced , for example, from 76 to 15 "*Issues*", by merging the lowest count issues into a single "*other issue*" category. Note that "*issue*" is correlated with "*product*".

C. Data size and distribution

The class distribution of complaints is highly skewed. For "*Product*" in this subset, "*Debt collection*" (17,552) accounts for 26.27% of the complaints, "*Money transfers"* (666) for about 1%, and "*Other financial service*" (110) for less than 1%. This is typical of real-world datasets. Fig. 3 is the histogram of complaints by "*Product*" category.

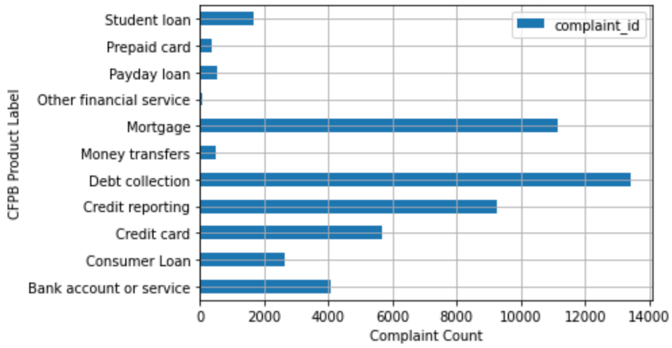

Fig. 3. CFPB Product Label Histogram

III. PROBLEM STATEMENT AND METHODOLOGY

A. Problem statement

Text classification is formulated as a supervised machine learning task. Given a collection $D = \{d_1, d_2, \ldots, d_N\}$ of N documents and a set $C = \{c_1, c_2, \ldots, c_K\}$ of K classes, the problem is to assign to each document d_j a corresponding class c_k in C.

In the *binary* classification case there are two possible classes (K=2). In the more general *multi-class* (also called *multinomial*) case, there are more than two classes (K>2). Similarly, the classification problem may be *single-label* (each document belongs to a single class) or *multi-label* (each document may belong to multiple classes).

The classification classes are the categories of the "*Product*" and "*Issue*" labels (11 products and 15 issues), and each consumer complaint is assigned to only one "*Product*" and one "*Issue*" category; Below, we consider only the "*Product*" label. Our problem is thus multi-class, single-label classification.

Additionally, in a supervised problem we have classification examples (d_1, c_1), (d_2, c_2), …, (d_T, c_T), T training examples from the dataset (supervised task), and our goal is to learn a *CCD* (Consumer Complaint Database) classifier from the training examples.

B. Methodology

In machine learning problems, such as text classification, the input and output data (input documents and output document classes) are represented as mathematical objects (feature vectors in a high-dimensional vector space), and the decision function, classifier or *model* to be learned maps the input representations to a corresponding output.

The quality of a model is measured probabilistically, in terms of loss or *error* the model makes in its classification decisions, and the goal is to learn a model that minimizes the error. Alternatively, the quality of a model and its evaluation may be couched in terms of *accuracy* of the decisions, and the goal is to maximize accuracy. For text classification, information retrieval metrics of *Precision*, *Recall* and *F1–measure*, are used as accuracy measures [2].

The consumer complaint documents may be viewed as sequences or characters or words over a given vocabulary (e.g, English), with significant structure imposed by language. However, a document may be usefully represented as set or "bag" of characters or words, or or as a set of *n-grams* (a short sequence of *n* characters or words).

A common document representation, used here, is the "*bag-of-words*" (BoW) representation, which may be binary, a normalized count, or a term frequency-weighted count of words in the document. This type of representation of documents looses the style and a lot of the meaning of a document, since word order is ignored. However, the representation has been highly successful for simpler document processing tasks like information retrieval and classification.

Assuming words are drawn from a vocabulary V, each document becomes a numerical vector of dimension $|V|$, where $|V|$ is the size of the vocabulary. Typically $|V|$ is tens of thousands to hundreds of thousands or words in a document collection. In this case, each document becomes a $|V|$-dimensional point in a document vector space (e.g., of dimension 20,000), and the goal of a classifier is to learn a function that separates documents in the vector space according to their class or classes (supervised machine learning).

As in statistics, model evaluation requires awareness and rigor in the use of data and data sampling methods for problem analysis, model development and validation, and model testing. In particular, an available working dataset may be split into training, validation, testing segments (e.g., 80% training, 10% validation, and 10% testing), following well-known practice.

Finally, data origin, vetting and ethics are increasingly important to consider, as solutions are considered, models are developed, and both are applied with important impacts on people and societies. These issues are briefly discussed in the final section of the paper.

C. Data preparation and representation

Data wrangling, exploratory data analysis and data manipulation are done with Python data science tools and libraries to discover issues with the data (e.g., missing values, duplicates, errors, etc.), determine the best feature representations of the data (feature engineering) and to transform the data to the formats needed by the machine learning models.

The key data science libraries for this work include NumPy, pandas [13], and Scikit-Learn [14], and deep learning libraries Tensorflow [14] and Keras [15].

For input preprocessing and representation in the first three models below, we restrict vocabulary size to $|V|$=20,000 (e.g., eliminating words that occur only once in the dataset) and use bag-of-words document vectorization with term-frequency, inverse document frequency (TF-IDF) weighting, using Scikit-Learn's TfidfVectorizer class. Each document is represented as a $|V|$-dimensional vector.

D. Previous work

Text classification [2, 3, 16] started with early work in information retrieval [11] and text analysis of documents with discrete symbolic representations of linguistic structure, including words, categories, phrases, meaning [17].

The practical development of statistical methods and large, trainable neural network models, since 2000, together with the success of deep learning since 2012 lead to new architectures, including word and character-level recurrent neural networks (RNNs), convolutional neural networks (CNNs) [7], long short-term memory (LSTM) [8], neural attention [9], and recently, transformers such as the BERT model, with Bidirectional Encoded Representations from Transformers [10].

Recent previous work on financial text analysis and the Consumer Financial Protection Bureau (CFPB) dataset can be found in [12, 18, 19].

IV. MACHINE LEARNING MODELS

We define two baseline text classification models and three modern deep learning models.

A. Baseline model 1: Bag-of-Words with Naive Bayes

Our first baseline model (BoW-MNB) is a Multinomial Naive Bayes classifier with the input preprocessing and representation described earlier, with bag-of-words, TF-IDF document vectorization.

```
# Model TfidfVectorizer
tokenizer = '(?u)\\b\\w\\w+\\b'
ngram = (1, 3)
max_features = 20000
vectorizer = TfidfVectorizer(
        token_pattern=tokenizer,
        ngram_range=ngram,
        max_features=max_features)

# Model: MNB
tfidf_mnb_model = make_pipeline(
        vectorizer, MultinomialNB())
```

This model has vocabulary size $|V|$=20,000, uses the Scikit-Learn's *MultinomialNB* and *TfidfVectorizer* classes, and reaches a 77.8% classification accuracy on test data.

B. Baseline model 2: Bag-of-Words with MLP

The second baseline model (BoW-MLP) is a multi-layer perceptron, with a similar BoW, TF-IDF document vectorization and vocabulary size $|V|$=20,000. The perceptron uses the Tensorflow-Keras API [14, 15].

```
# BoW-MLP model
output_classes = 11 # dataset classes
dense_nodes = 128    # hyperparam
dropout = 0.5        # hyperparam

model = Sequential()
model.add(Dense(dense_nodes,
        activation='relu', input_shape=(20000,)))
model.add(Dropout(dropout))
model.add(Dense(dense_nodes, activation='relu'))
model.add(Dense(
        output_classes, activation='softmax'))
```

This model has vocabulary size $|V|$=20,000, uses the Scikit-Learn and *TensorFlow/Keras* deep learning APIs, and reaches a 84.4% classification accuracy on test data.

C. Bidirectional LSTM (Bi-LSTM) model

The next model is a word-based long short-term memory [11], with a trainable input-word embeddings layer of embedding dimension 100 (E100-Bi-LSTM).

```
# E100-Bi-LSTM model
lstm_size = 64        # LSTM hidden nodes
embed_dim = 100       # hyperparam
dropout = 0.3         # hyperparam

model = tf.keras.models.Sequential([
    tf.keras.layers.Embedding(
        max_features, embed_dim),
    tf.keras.layers.Bidirectional(
        tf.keras.layers.LSTM(
        lstm_size, dropout=dropout,
        recurrent_dropout=dropout)),
    tf.keras.layers.Dense(
        dense_nodes, activation='relu'),
    tf.keras.layers.Dropout(dropout),
    tf.keras.layers.Dense(
        output_classes, activation='softmax')
])
```

This model has similar vocabulary size $|V|$=20,000, and also uses the *TensorFlow*/*Keras* deep learning APIs. It reaches reaches a 83.1% classification accuracy on test data. Fig. 4. shows the Bi-LSTM model training over 10 epochs.

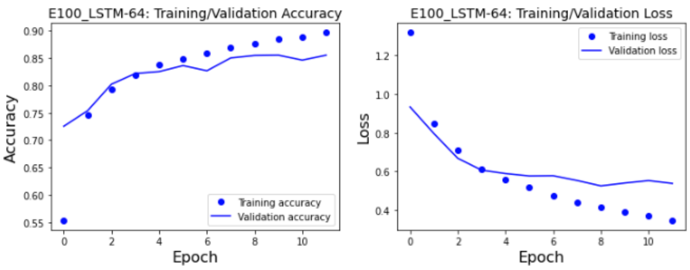

Fig. 4. Bi-LSTM Train/Dev loss and accuracy

Our final two models implement neural attention and are based on the BERT language model [10].

D. DistilBERT: Transformer large language model

DistilBERT is a distilled version of BERT [20], with 40% fewer parameters and yet comparable performance. It is found in the HuggingFace transformers library [21]. It is a pre-trained large language model (LLM) with 66,961,931 parameters [DistilBERT-base-uncased].

```
# DistilBERT transformer model
model_ckpt = "distilbert-base-uncased"
num_labels = len(cfpb_product_names)
batch_size = 32

tf_model =
hf.TFAutoModelForSequenceClassification.
from_pretrained(model_ckpt, num_labels=num_labels)

tf_model.compile(optimizer=tf.keras.optimizers.
Adam(learning_rate=5e-5),
loss=tf.keras.losses.
SparseCategoricalCrossentropy(from_logits=True),
metrics=tf.metrics.SparseCategoricalAccuracy())
```

DistilBERT is fine-tuned end-to-end for two epochs with batch size 32 on CFPB training data and reaches an 87.05% classification accuracy on test data.

E. FinBERT: Transformer large language model

Our final model is FinBERT, a BERT variant pre-trained for sentiment analysis on financial texts [22]. It is also a pre-trained large language model with 109,490,699 parameters [ProsusAI/finbert]. The code is the same as for DistilBERT, except for the different model checkpoint.

```
# FinBERT transformer model
model_ckpt = "ProsusAI/finbert"
```

The pre-trained FinBERT is fine-tuned end-to-end for two epochs on CFPB training data and reaches an 88.05% classification accuracy on test data.

In Fig. 5, the classification report shows the precision, recall and F-1 accuracy measure of the model for each of the output classification categories.

	precision	recall	f1-score	support
Bank account or service	0.91	0.81	0.85	193
Consumer Loan	0.90	0.81	0.85	173
Credit card	0.86	0.89	0.87	285
Credit reporting	0.88	0.96	0.92	363
Debt collection	0.87	0.79	0.83	400
Money transfers	0.70	0.54	0.61	13
Mortgage	0.92	0.97	0.94	474
Other financial service	0.00	0.00	0.00	5
Payday loan	0.69	0.44	0.54	25
Prepaid card	0.61	0.94	0.74	18
Student loan	0.83	0.96	0.89	51
accuracy			0.88	2000
macro avg	0.74	0.74	0.73	2000
weighted avg	0.88	0.88	0.88	2000

Fig. 5. FinBERT Classification Report on CFPB Data

In Fig. 6, the classification performance of the model is visualized in the model confusion matrix on the development or test sets, comparing expected model labels (ground truth) to the actual predicted labels from the model.

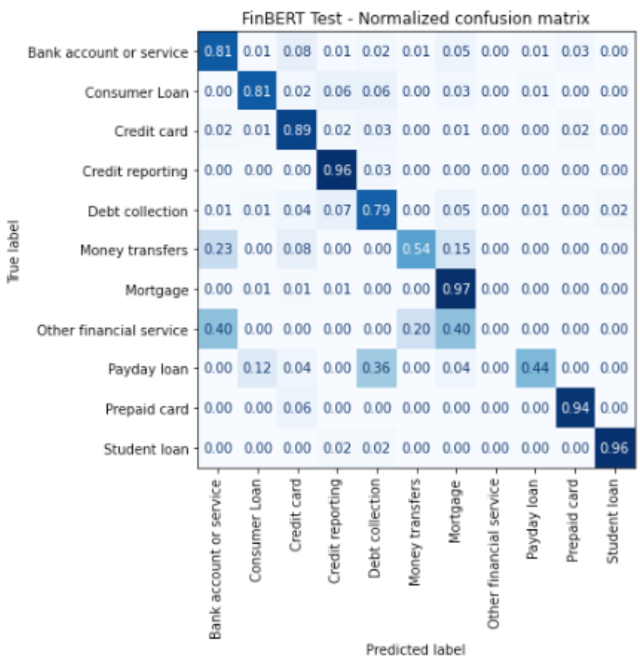

FIG. 6. FinBERT Test Confusion Matrix

A GitHub repository with models, python code, dataset, and Jupyter notebooks is available.[1]

V. Metrics and Evaluation

A. Hyper-parameters and classification accuracy

We trained transformer models for two epochs, with batch size 32 and default model hyper-parameters. The best model, ProsusAI/FinBERT, has an architecture BertForSequenceClassification, with parameters as follow.

```
# FinBERT model hyper-parameters

"attention_probs_dropout_prob": 0.1,
"hidden_dropout_prob": 0.1,
"hidden_size": 768,
"layer_norm_eps": 1e-12,
"max_position_embeddings": 512,
"num_attention_heads": 12,
"num_hidden_layers": 12,
"vocab_size": 30522
```

We expect model performance to increase with more extensive training (additional CFPB data and more epochs) and hyper-parameter tuning.

B. F1 and other evaluation metrics

Table 1 summarizes the model performance of the five models considered here: For each model we show the number of parameters of the model, and its accuracy on the test and validation splits.

Model	Model Parameters	Test Accuracy	Validation Accuracy
BoW-MNB	--	77.8%	79.0%
BoW-MLP	2,577,801	84.4%	86.7%
E100-Bi-LSTM	2,093,451	83.1%	85.6%
Fine-tuned DistilBERT-base	66,961,931	87.05%	86.86%
Fine-tuned ProsusAI/FinBERT	109,490,699	88.05%	87.56%

Table 1. Model parameters and Dev-Test accuracy

The best accuracies obtained, above 85%, are competitive with human levels of performance, as can be inferred from human inter-annotator agreement on similar multi-class classification tasks [2].

Note that the classification accuracy of a simple baseline classifier would be 9.09% (1/11) if a random class out of eleven is selected for each document, or 26.27% (17,552/66,806) if instead the most likely class ("*Debt collection*") is always selected as a baseline. All models significantly outperform these baselines. The transformer models outperform other models by an absolute 4% points.

C. Data size and compute equipment

Exploratory data analysis of the full CFPB dataset of over 2,600,000 complaints in JSON format was done on an 8-core Ubuntu notebook. The text classification models and results presented here use the Kaggle CFPB dataset with 555,957 records and 66,806 with non-empty consumer complaint narrative. Data analysis, models and results were run on a high-end MacBook PRO.

D. Societal and ethical considerations

The development of artificial intelligence and natural language processing applications poses societal and ethical implications for which there needs to be awareness, and that must be monitored and evaluated. Those include questions of (1) origins of the data and representativeness of populations impacted by it, and (2) questions of explainability, bias, fairness, future of work and other societal consequences that use of the data and development of artificial intelligence poses.

Regarding (1), the demographic distribution of users represented in the data (ethnic origin, race, gender, and literacy level among others) may impact the quality of the applications (ML models) that may be created from it, the applications being highly accurate and thus beneficial for one group of users, but not always for others.

Regarding (2), the data and methods used for model development and validation may similarly impact the quality of applications for users, and the fairness and bias that applications may show to one group of users. Increasingly, it is important that machine learning models be interpretable and explainable in terms of the decisions they make [23].

One component to address these issues is a *dataset and model card* for each dataset and ML model created. A model card documents the machine learning model with information about the training algorithms and parameters; training data sources, motivation, and preprocessing; evaluation data sources, motivation, and preprocessing; intended use and users; and model performance across different demographic or other groups and environmental situations [24].

This is an evolving issue, and governments and large organizations increasingly pay attention to it, including the European Union [25], the United States [26], and several academic institutions working in the field [27]. There are already concrete regulations and proposals for the development and deployment of AI/ML applications, including that presented here. Thus, organizations look to AI/ML Ethics and Governance panels, to develop an inventory and track the use of the new technologies.

VI. Conclusion

Text classification and other natural language processing tasks are increasingly important use cases for business process automation in business and government, enabling the efficient and reliable classification, routing and other uses of billions of documents. We described the CFPB Consumer Complaints Database and implemented classical and recent large language machine learning models for multi-class complaint classification, obtaining an accuracy of 88.05% with the FinBERT transformer model. This accuracy is competitive with the human level of performance on similar multi-class classification tasks, and the models developed are thus suitable for practical document automation and other business solutions. Human performance evaluation on our datasets and tasks is an effort that we explore in future work.

Acknowledgment

We thank Christine Braun-Thurnher, Pablo Duboue, Gabriela Correa, K.P. Unnikrishnan, Wlodek Zadrozny and our three reviewers for useful discussions, and editorial and technical comments that helped to improve this paper.

[1] *https://github.com/nelscorrea/andescon2022*

References

All cited online references were accessed between January and April, 2022.

1. Stuart Russell and Peter Norvig, 2020, Artificial Intelligence: A Modern Approach, 4th Edition. Pearson.
2. Daniel Jurafsky and James Martin, 2022, Speech and Language Processing, 3rd Edition (draft), https://web.stanford.edu/~jurafsky/slp3/.
3. Bing Liu, 2020, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge.
4. Thomas Siebel, 2019, Digital Transformation: Survive and Thrive in an Era of Mass Extinction, Rosetta Books.
5. Consumer Financial Protection Bureau, Consumer complaints database: https://www.consumerfinance.gov/data-research/consumer-complaints/
6. Dodd-Frank Act, 2010, Dodd-Frank Wall Street Reform and Consumer Protection Act, Public Law 111–203, https://www.govinfo.gov/content/pkg/COMPS-9515/pdf/COMPS-9515.pdf.
7. Xiang Zhang, Junbo Zhao, Yann LeCun, 2015, Character-level Convolutional Networks for Text Classification, NIPS 2015.
8. Sepp Hochreiter and Jürgen Schmidhuber, 1997, Long Short-Term Memory, Neural Computation, Vol. 9, Issue 8, https://doi.org/10.1162/neco.1997.9.8.1735.
9. Dzmitry Bahdanau, K.Y. Cho, and Y. Bengio, 2015, Neural Machine Translation by Jointly Learning to Align and Translate, ICLR 2015, https://arxiv.org/abs/1409.0473.
10. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, 2018, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2019: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1.
11. Peter Jackson and Isabelle Mulinier, 2007, Natural Language Processing for Online Applications: Text retrieval, extraction and categorization, John Benjamins B.V.
12. Kaggle, 2016, US Consumer Finance Complaints: US consumer complaints on financial products and company responses. https://www.kaggle.com/datasets/kaggle/us-consumer-finance-complaints.
13. Wes McKinney, 2022, Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, Second Edition, O'Reilly.
14. Aurélien Géron, 2019, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, Second Edition, O'Reilly.
15. François Chollet, 2021, Deep Learning with Python, Second Edition, O'Reilly.
16. Bo Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. EMNLP 2002.
17. Nelson Correa, 1991, Empty Categories, Chain Binding, and Parsing. in *Principle-based Parsing: Computation and Psycholinguistics*, R. C. Berwick, C. Tenny, and S. Abney, eds. Kluwer Academic Publishing, Dordrecht, Netherlands.
18. Vasudeva Raju Sangaraju, Bharath Kumar Bolla, Deepak Kumar Nayak, Jyothsna Kh, 2022, Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings, https://arxiv.org/abs/2205.07259.
19. Bernhard Lutz, Nicolas Prollochs, Dirk Neumann, 2018, Sentence-Level Sentiment Analysis of Financial News Using Distributed Text Representations and Multi-Instance Learning, https://arxiv.org/abs/1901.00400.
20. Victor Sanh, L. Debut, J. Chaumond, T. Wolf, 2019, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, https://arxiv.org/abs/1910.01108.
21. Lewis Tunstall, L. von Werra, T. Wolf, 2022, Natural Language Processing with Transformers, O'Reilly.
22. Dogu Araci, 2019, FinBERT: Financial Sentiment Analysis with Pre-trained Language Models, University of Amsterdam, June 25, 2019. https://arxiv.org/abs/1908.10063.
23. Nelson Correa, 2020, NLP and Artificial Intelligence: Neural and Symbolic Approaches, Florida Atlantic University, Data Science, Analytics, and AI Conference Boca Raton, Florida, 2020, https://www.fau.edu/data/schedule/2020/.
24. Melanie Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru. 2019. Model cards for model reporting. ACM FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency. ACM.
25. European Commission (https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence)
26. U.S. National Artificial Intelligence Initiative (https://www.ai.gov)
27. Stanford Human-Centered AI (https://hai.stanford.edu)