Language Detection Using NLP

[1st] Aadersh Jaiswal
School of Computer Science and Engineering, Galgotis University, Greater Noida, India
aadrshjaiswal84@gmail.com

[2nd] Aahana Mahto
School of Computer Science and Engineering, Galgotis University, Greater Noida, India
ahanamahto795@gmail.com

[3rd] Satyam Kumar Singh
School of Computer Science and Engineering, Galgotis University, Greater Noida, India
satyam8511322@gmail.com

[4th] Raju Kumar
School of Computer Science and Engineering, Galgotis University, Greater Noida, India
rajuk12@gmail.com

[5th] Rydhm Beri
Alliance College of Engineering and Design, Alliance University, Bengaluru, India
rydhmberi@gmail.com

[6th] Kajal Rai Saraswat
Department of MCA, GL Bajaj Institute of Technology and Management, Greater Noida, India
kajalrai.pu@gmail.com

Abstract*— A technique for accurately detecting text based on the given content or topic matter is called natural language processing, or NLP. It will be easy to read any language and understand what is being stated with a thorough study. NLP is a difficult method, however there are some prominent examples, including Siri and Alexa. We can ascertain the language being utilised in a particular document by using natural language detection. The model utilised in this work, which was created in Python, can be used to analyse any language's fundamental linguistics. The fundamental units of knowledge and its expression are the "words" that comprise sentences. It is crucial to recognise them correctly and understand the context in which they are employed. NLP intervenes to assist us in this by making it simpler for us to recognise the language employed in a specific informational item, whether it be spoken or written. NLP enables computers to identify language for us by understanding human speech and responding appropriately. This document summarises the progress made in the linguistic process, encompassing establishment, analysis, a number of quickly expanding fields of natural language processing research, development tools, and approaches.

Keywords: natural language processing, text analytics, virtual assistants, machine learning, and language detection

I. INTRODUCTION

The process of processing languages and converting them into forms that users can easily understand or utilise is known as natural language processing, or NLP. Pattern recognition is the foundation of computer programming that uses natural language processing (NLP) [1]. Natural Language Generation (NLG) and Natural Language Understanding (NLU) are its two components. NLU can be used to ascertain the meaning of a text passage, regardless of whether it is written or spoken. NLG generates meaningful sentences based on a representation of text or data. NLP is the foundation upon which Language Detection operates. Language recognition and processing are done with NLP. Various word and language varieties can be identified with the help of NLP. NLP helps with text analysis by recognising words and languages. Interpretation. NLP facilitates the recognition of commercial writings. NLP helps us implement and detect many languages by identifying the databases to whose language each speaks and interpreting the text to determine its purpose and meaning. using a number of libraries and datasets, NLP can be used to accomplish the same goal with greater assistance and coverage. Because most NLP applications are language specific, they require monolingual data. Preprocessing and removing content published To construct an application in the target language, proficiency in languages other than the target language may be necessary [2]. For example, we need to specify the exact language of each input. The processes involved in processing natural language include discourse synthesis, syntactic analysis, semantic analysis, lexical (or structural) analysis, and pragmatic analysis. Scanners, speech detectors, text chats, and computational linguistics are among the common applications of language technology. In order to operate tongue words, we now analyse large employing artificial intelligence (AI) approaches, samples of human-written words (conversation, keywords, and details) are produced [3]. Computers can be trained to understand the "context" of writing, human speech, and other human communication through the examination of these patterns. Deep learning and machine learning algorithms are widely utilised to create NLP frameworks and effectively finish common NLP jobs [1]. Utilising natural language processing Language processing is now growing at a very rapid rate.

II. LITERATURE REVIEW

While syntactic structures and its rule-based system, including the "Turing Test," were developed in 1950 and 1957, respectively, the original beginnings of NLP research took place in the late 1940s. Up until 1990, growth was sluggish because of a lack of computer capability, the usage of manual rule-based systems, and a limited vocabulary. Due to advancements in the discipline and the continuous increase in computing capacity, interest in machine learning has lately increased [15]. Among the most significant NLP breakthrough areas in recent years include language processing, speech recognition, dialogue systems, and deep learning techniques. Researchers have shown a

great deal of interest in NLP. and created numerous chances to apply its methods to robotics, automation, and digital transformation. Notwithstanding the difficulties it continues to encounter (such as those pertaining to HCIs) [3]. Most of the research on machine translation and NLP principles was completed before 1990. The most recent NLP research has made extensive use of computational learning, deep learning, and statistical models. There are times when studies in natural language processing and artificial intelligence—particularly deep learning—converge. These days, NLP jobs are typically completed as efficiently as feasible by using these strategies [1]. Speaking with a machine will eventually be just as easy as speaking with a human. Unstructured data is still used by NLP to provide meaning for a machine. Robotics, healthcare, finance, connected cars, among the sectors that stand to benefit from NLP are and smart homes [2]. Natural language processing (NLP) was first used in machine translation, which translated words from one human language to another. 21st century[13]. Nonetheless, it gained popularity in the customer service sector right away. One of the most popular NLP customer service tools is the virtual assistant, or "Chatbot." Different sectors employ different applications. Below is a list of these:

A Conversational systems

A conversational system enables humans to converse through a speech or text interaction with an automated system that speaks in natural language [2]. They help companies automate difficult jobs. so they can provide 24/7 customer support. Chatbots and virtual assistants are the two types of conversational technologies that are most frequently used. These days, self-service point-of-sale systems, social media, banking, and e- commerce all use then two gadgets to provide its clients a variety of services.

B Text Analytics

Tex analytics, also often known as text mining, seeks to extract significant information from text, regardless of text length, ranging from lengthier documents and emails to shorter messages like tweets and SMS texts [23]. One of the most popular uses of text analytics is social media analysis.

C Machine Translation in

Preserving the intended The goal of machine translation is to automatically translate text between natural languages while maintaining meaning. Google Translate is the most widely used machine translation application. Additional machine translation software is also used in speech translation and education [14]. NLP is also used in the manufacturing, healthcare, retail, automotive, financial, and educational sectors. Hospitals are using virtual assistants that were created by fusing natural language processing, machine learning, and computer vision. Based on their contacts with patients, these virtual assistants will automatically create and assemble patient histories [12][25]. Virtual assistants oversee daily operations, such as patient registration and organising appointments.

One of the most exciting new advancements in the industrial sector is the creation of self-driving cars. which NLP makes possible and are growing in acceptance within the sector. NLP-based technologies are utilised in the banking industry to develop applications including credit scoring, document search, and sentiment analysis. Through the use of NLP and machine learning, credit scoring algorithms enable credit agencies, banks and other lending organisations to assess an individual's creditworthiness and provide a credit score. Sentiment analysis applications automate named entity recognition and document classification processes to choose the data most pertinent to investor needs Chatbots are used in banks and other financial institutions. in document search apps to enable their customers to search for information and receive straightforward transactional responses [24].

Two really promising NLP application areas are robotics and process automation. A robot processes instructions to assemble and move machinery and items.

on a production line can conversewith natural language processing (NLP) being used by a human operator situated remotely [4].

A retail virtual assistant may recognise and understand a customer's needs when it is in front of a retail establishment and provide them with timely information and promotional offers by utilising Natural Language, Computer Vision, and Machine Learning technologies [10].

The integration of computer vision and natural language processing in educational platforms allows for the provision of a virtual classroom for pupils.

The use of digital assistants to help students with specialised knowledge from online libraries has already proven beneficial [9].

D NLP Progress: Frameworks and Tools Today's development tools are widely available because open-source communities worldwide have shown such a strong interest in them [6]. These tools and frameworks have integrated functionality and might be altered to adhere to particular industrial norms. shelves for books.

Graph, tree, or structured representations are used by the natural language representation block to express natural language information [7]. A Natural Language database, akin to MNIST or other databases, is an assemblage of Natural Additional NLP is performed by machine learning algorithms using language information. tasks.

The representation and transformation blocks use this database to carry out their operations. A variety of learning and extraction strategies will be used in natural language transformation to extract relevant and meaningful tasks from NLP jobs [5]. The presentation of natural language communication is the the actions that tasks using natural language processing (NLP) are intended and desired to produce [11]. Either natural

language or computer activity, such as a robot arm moving, could be the outcome.

Human conversation has led to the development of natural language processing. Without a doubt, the process will entail translating natural human language into a format that can be understood by machines. NLP tasks could include the following tasks:

1) Word sense ambiguation is the process of choosing a word from among several meanings by using semantic analysis to determine which word is most appropriate in a given situation.

2) Speech Recognition: This method transforms audio inputinto data in text format.

3) Named Entity Recognition: This innovation in technology recognises words as pertinent and helpful entities.

4) Part of speech tagging: This method ascertains the part of speech of a specific text in a sentence or information by utilising the best context that is available.

Natural language generation (NLG) and natural language comprehension (NLU) are the two components of NLP.

. NLU: It includes the subsequent:- a. Lexical ambiguity: This occurs when a word's appropriate and pertinent meaning needs to be determined inside a text.

b. Referential ambiguity: This occurs when a term appears more than once in a phrase.

d. Syntactical Ambiguity: Perceiving many meanings within a text.

NLG: This method involves translating structured data into understandable language for humans [20]. It takes a transforms the translation of data or material into coherent sentences. Sentence planning, which entails selecting suitable words and phrases for a written piece, is one of its components.

b. Text Planning: This provides us with pertinent data and statistics. from an information repository.

c. Text Realisation: This is how sentence structure and sentence plan are mapped.

Sentiment analysis, a method Natural language processing also includes an element that use statistical analysis to determine the intent and meaning of the emotionally charged content.

As a subset of NLP, Language Detection (LD) is included. As was previously mentioned, it is based on the NLP principle [19]. In this case, the language and linguistics employed in a certain written work or knowledge base are assessed and identified in their format. Here, the language in which the content is found is identified [11]. This topic is approached computationally as a specific case of text categorization, which is resolved using a variety of statistical techniques [21]. LD is a fantastic method for quickly and effectively classifying information, sorting it, and applying extra language-specific workflow layers [22]. It can assist us in recognising and detecting spelling or grammar mistakes in a specific document. For instance, let's say we compose a statement in English with a specific spelling mistake [18]. The system may then assist us in analysing the text and identifying the language that is written as "English" by employing the principle of Language Detection to help us find and fix spelling mistakes in words that are written improperly. NLP contains numerous libraries, including genism, spaCy, and NLTK [16]. These repositories

support the creation of NLP models and the application of NLP traits. These accomplish their goal by providing a great deal of assistance to Language Detection models.

Section III: Methodology "Google Colab" platforms are used for implementation. A prepared "Language Detection Using NLP" file is used to load the data. A dataset obtained from Github and Kaggle is utilised to train a model. Based on the requirements, just a select The downloaded dataset contained several languages; only a few of them were chosen. We will thoroughly discuss each implementation.

- Step 1: Importing all of the libraries and packages required to complete the task is the first step. For example, mounting a dataset from a local computer to Google Drive is the second step. the following action plan. The dataset is stored as a zip file on Google Drive in the cloud.At the moment, the dataset is mounted on Google Drive in the "Google Colab" environment. About 80 GB of local storage are available to us on the Google Colab Environment's distributed server.

- Step 3: A CSV file's data can be retrieved as a data frame by employing the read_csv() method.

- Step 4: At this point, we'll identify the crucial variable. that is crucial to the development of our machine learning model. The variable names and corresponding values are shown in the illustration.

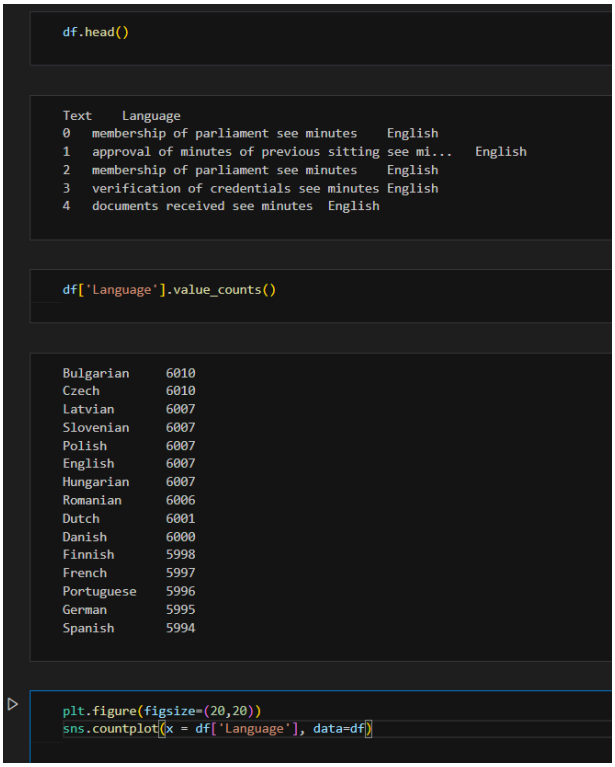


Fig 1. Defining the Variable

The variables in Fig. 1 have the following definitions:

• head

The head method in Python by default displays the first five rows of the data frame. The number of rows is the sole parameter that it takes into account. This reasoning enables us to recognise the number of rows we want to see.To obtain the dataframe's first n rows, use head(n). It takes one optional argument, n, which is the desired initial row count.

• Value_count ()

Value_counts is a function that yields an object containing counts of unique values. This will lead to the appearance of the object in descending order, with the most frequent element appearing first.

• Step 5: The This uses the class label encoder from the sklearn package; a description of its entire usage follows below:

The category feature levels are encoded into numerical values by Sklearn, which provides an incredibly powerful tool. In order to transform the labels into a format that a machine can read, a process known as label encoding must be performed. After them, machine learning algorithms will be able to make better decisions about how to use the labels. It's an crucial stage in the structured dataset pre-processing process for supervised learning. Labels ranging in value from 0 to n_classes-1, where n is the total number of distinct labels, can be encoded by LabelEncoder. If a label appears more than once, the value that was previously assigned is used.

Using the fit_transform() method, Fitting the label encoder converts multi-class labels into binary labels. The coding system 1-of-K is another name for the result of this conversion. Fig. 2 shows Fit_Trasform and LableEncoder included.

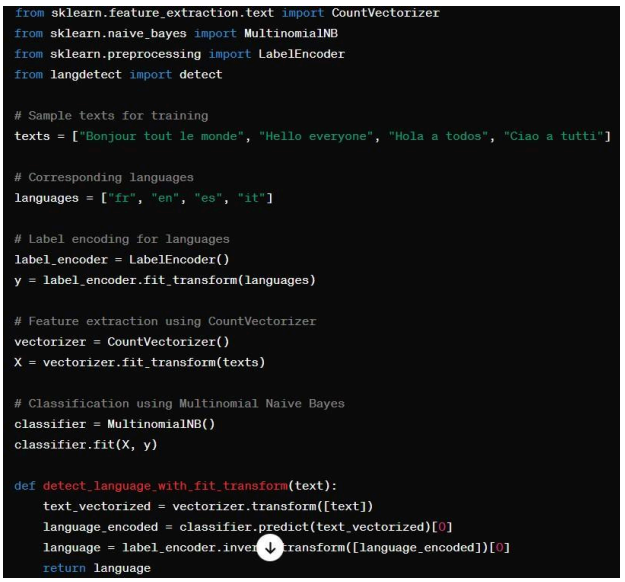


Fig. 2: Fit_Trasform with LableEncoder included

• Step 6:

A new array called data_list is made, and the Python Regular Expressions (re) module's re.sub() function is utilised. It will return a string in which the replace string will replace every instance that matches the given pattern. The function re.sub() takes a substring as input and returns a string with its values changed. The.lower function is used to transform all letters to lower case, and it allows us to replace multiple items by using a list.

After that, the append function is used to add the text to the data_list array.

The following usage criteria apply to the sklearn.feature_extraction module class, which is part of the sklearn module and will be used in this instance: To extract features in a machine learning-supported and protected format methods, the sklearn.feature_extraction module is used. Typically, this module is used with datasets that contain text and image formats.

Utilising a method provided by the Python scikit-learn package, CountVectorizer is a helpful tool. Using it, one can apply the transformation of a text into a vector according to each word's frequency (count) or appearance in the text [8]. This may be really beneficial if we want to convert every word in several texts into a vector so we can utilise it in a future text analysis.

By using A matrix with a matrix column is created using CountVectorizer. a distinct image for every word, and each row in the matrix corresponds to a text excerpt from that specific source. The number of words in the specific text sample provided determines the value of each individual cell.

We will employ the fit_transform () function, which is essentially the same as transform().fit() but combines the transform and fit methods. This approach converts data points by performing a fit and change the concurrently input data.

Next, we set the dimension of the array X using the.shape method. An array data_list is generated in Fig. 3.

```
# Cleaning the data for english dataset

data_eng = []
lang_eng = []

for i, line in english_df.iterrows():
    line = line['English']
    if len(line) !=0:
        line = line.lower()
        line = re.sub(r"\d+", "", line)
        line = line.translate(translate_table)
        data_eng.append(line)
        lang_eng.append("English")
# Cleaning the data for german dataset

data_ger = []
lang_ger = []

for i, line in german_df.iterrows():
    line = line['German']
    if len(line) !=0:
        line = line.lower()
        line = re.sub(r"\d+", "", line)
        line = line.translate(translate_table)
        data_ger.append(line)
        lang_ger.append("German")
# Cleaning the data for french dataset

data_fre = []
lang_fre = []

for i, line in french_df.iterrows():
    line = line['French']
    if len(line) !=0:
        line = line.lower()
        line = re.sub(r"\d+", "", line)
        line = line.translate(translate_table)
        data_fre.append(line)
        lang_fre.append("French")
```

```
        # Cleaning the data for Bulgarian dataset

data_bul = []
lang_bul = []

for i, line in bulgarian_df.iterrows():
    line = line['Bulgarian']
    if len(line) !=0:
        line = line.lower()
        line = re.sub(r"\d+", "", line)
        line = line.translate(translate_table)
        data_esp.append(line)
        lang_esp.append("Bulgarian")
# Cleaning the data for Czech dataset

data_czech = []
lang_czech = []

for i, line in czech_df.iterrows():
    line = line['Czech']
    if len(line) !=0:
        line = line.lower()
        line = re.sub(r"\d+", "", line)
        line = line.translate(translate_table)
        data_esp.append(line)
        lang_esp.append("Czech")
# Cleaning the data for Danish dataset

data_dan = []
lang_dan = []

for i, line in danish_df.iterrows():
    line = line['Danish']
    if len(line) !=0:
        line = line.lower()
        line = re.sub(r"\d+", "", line)
        line = line.translate(translate_table)
        data_esp.append(line)
        lang_esp.append("Danish")
```

Fig: 3 Creating an array data_list

STEP 7: The list is divided into a training set and a testing set by this section of code. It's a fundamental concept in the building of models for machine learning [17]. The sklearn module's train test split method makes use of the test_size, X, and Y parameters. The separation of the Fig. 4 displays the dataset split into training and testing datasets..

The following is a list of the four variables that comprise the training and testing dataset:

1. X_train
2. Y_train
3. X_test
4.Y_test

```
from sklearn.model_selection import train_test_split

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(texts, languages, test_size=0.2, random_state=42)
```

Fig 4: Separating the dataset into datasets for testing and training

The MultinomiaNb module's model is found in Step VIII. This step involves building a neural network model using fit(). This is an extra useful Naïve Bayes classifier. This makes the assumption the features are drawn from a basic multinomial distribution. Scikit-learn provides Neave Bayes at sklearn. Implementing the Multinomial Naïve Bayes classification technique is possible with MultinomialNB. We are now using MultinomalNB's fit technique, which takes x and y as input. At this point, the target values, or y, should be represented by y, and the training vectors, or training data, by x. Figure 5 shows the creation of a neural network model.

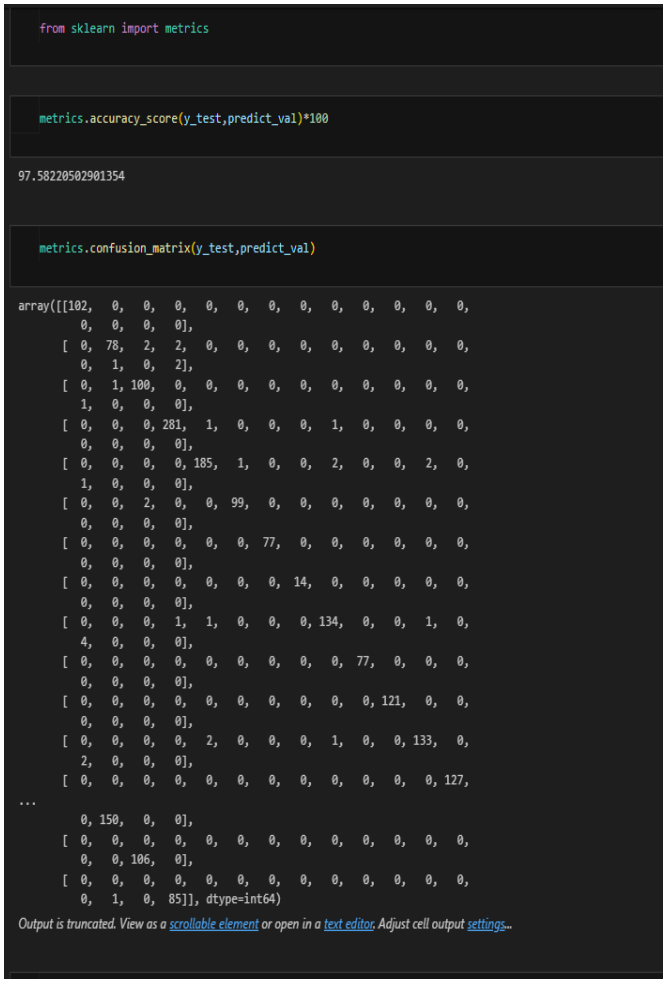
```
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB()
model.fit(x_train,y_train)
MultinomialNB()
```

Figure 5: Constructing a Neural Network Model

- phase-8: The correctness of the model is determined in this phase. In Fig. 6, the model accuracy is displayed.

```
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
ac = accuracy_score(y_test,y_pred)
cm = confusion_matrix(y_test,y_pred)
cr = classification_report(y_test,y_pred)
```

Find the Accuracy of the Model in Fig. 6



```
from sklearn import metrics

metrics.accuracy_score(y_test,predict_val)*100

97.58220502901354

metrics.confusion_matrix(y_test,predict_val)

array([[102,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,
          0,   0,   0,   0],
       [  0,  78,   2,   2,   0,   0,   0,   0,   0,   0,   0,   0,
          0,   1,   0,   2],
       [  0,   1, 100,   0,   0,   0,   0,   0,   0,   0,   0,   0,
          1,   0,   0,   0],
       [  0,   0,   0, 281,   1,   0,   0,   0,   1,   0,   0,   0,
          0,   0,   0,   0],
       [  0,   0,   0,   0, 185,   1,   0,   0,   2,   0,   0,   2,   0,
          1,   0,   0,   0],
       [  0,   0,   2,   0,   0,  99,   0,   0,   0,   0,   0,   0,   0,
          0,   0,   0,   0],
       [  0,   0,   0,   0,   0,   0,  77,   0,   0,   0,   0,   0,   0,
          0,   0,   0,   0],
       [  0,   0,   0,   0,   0,   0,   0,  14,   0,   0,   0,   0,   0,
          0,   0,   0,   0],
       [  0,   0,   0,   1,   1,   0,   0,   0, 134,   0,   0,   1,   0,
          4,   0,   0,   0],
       [  0,   0,   0,   0,   0,   0,   0,   0,   0,  77,   0,   0,   0,
          0,   0,   0,   0],
       [  0,   0,   0,   0,   0,   0,   0,   0,   0,   0, 121,   0,   0,
          0,   0,   0,   0],
       [  0,   0,   0,   0,   2,   0,   0,   0,   1,   0,   0, 133,   0,
          2,   0,   0,   0],
       [  0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0, 127,
...
          0, 150,   0,   0],
       [  0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,
          0,   0, 106,   0],
       [  0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,
          0,   1,   0,  85]], dtype=int64)
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Fig 7: The model's accuracy

III. RESULTS

In this experiment, an overall accuracy of 0.98 was achieved. The model can be regarded as an excellent fit for this kind of study given its 98% accuracy.

IV. CONCLUSION

The modern world's growing reliance on technology has also led to a need for greater standards, which serve to validate the daily advancements we witness. Here, natural language processing and language detection lead to larger and broader fields that can facilitate human task completion by assisting in the much easier, better, and systematic recognition of texts; consequently, using statistical methods, these fields can facilitate technical work for people. As a result, we have made an effort to develop a Language Detection model using Natural Language Processing, which can help us properly and efficiently with Language Detection problems quickly recognising text using suitable techniques.

References

[1] In July 2018, Daniel W. Otter, Julian R. Medina, and Jugal K. Kalita published a 35-page article in Volume 1, Issue 1. An Overview of Deep Learning's Use in Natural Language Processing

[2] Robert Dale. Natural Language Engineering article "The Commercial NLP Landscape in 2017" published in July 2017.

[3] ACL 2018: The Association for Computational Linguistics' 56th Annual Meeting

[4] Predictive Analytics Today's website is located at www.predictiveanalyticstoday.com.[retrieved December 2018]

[5] Mahmoud Al-Ayyoub, Ghadeer Al-Bdour, Raffi Al-Qurran, and Ali Shatnawi 2018. An Analysis of Open Source Deep Learning Frameworks in Comparison. 2018: 9th International Conference on Information and Communication Systems (ICICS)

[6] EY study, 2018, Intelligent automation: Making cognitive real Knowledge Series I Chapter 2.

[7] Jacques Bughin, TeraAllas, Peter Dahlström, Eric Hazan, Sreeramaswamy, Michael Chui, Nicolaus Henke, Trench, Monica (2017). A NEW ERA IN DIGITAL INTELLIGENCE: MGI ARTIFICIAL INTELLIGENCE? McKinsey & Company July 2017 McKinsey & Company report

[8] Sicular Svetlana, Brant Kenneth 2018 Artificial Intelligence Hype Cycle, 2018, Gartner research, July 2018.

[9] Funda Durupinar, Oshin Agarwal, Ani Nenkova, and Norman I. Badler. 2019. Phrase embeddings also encode human personality stereotypes. Proceedings of the Joint Conference on Lexical and Computational Semantics, Minneapolis, MN, pages 205–211.

[10] Quarteroni, Silvia (2018). ELCA's Industrial Experience with Natural Language Processing. The Scientific Spectrum 41.10.1007/s00287-018-1094-1.

[11] Erik Cambria, Tom Young, Poia, Soujanya, Hazarika, and Devamanyu. (2018). Recent Developments in Deep Learning for Natural Language Processing [Review Article]. 13.55-75.10.1109/MCI.2018.2840738 IEEE Journal on Computational Intelligence.

[12] Chris Chandler, Hassan Kazemian, Ouazzane, Karim, and Mohammad Hossein Amirhosseini (2018) an automated method utilising natural language processing for NLP meta models. Proceedings of the 10th International Joint Conference on Neural Networks (IJCNN), July 8–13, 2018, Rio de Janeiro, Brazil.

[13] 2020; Barbara Plank & Alan Ramponi. An overview of natural language processing's neural unsupervised domain adaptability. Documents from the 28th International

[16] Abzaliev, Artem. 2019. Regarding the shared task of GAP coreference resolution: insights from the third-place solution.Pages 107–112, Florence, Italy: Proceedings of the Workshop on Gender Bias in Natural Language Processing.

[17] Zihang Dai, Eduard Hovy, Quoc Le, Thang Luong, and Qizhe Xie. 2020. Unsupervised

enhancement of data for training consistency. Neural Information Processing Systems Advances, 33

[18] Preetam Amancharla, Anupam Datta, Fangjing Wu, Piotr Mardziel, and Kaiji Lu. 2020. In neural natural language processing, gender bias pages 189–
202. International Publishing Springer, Cham.Miller, George A. (1995). Wordnet is an English lexical database. ACM Communications, 38(11): 39–41.

[19] Hal Daume III, Su Lin Blodgett, Solon Barocas, and Hannah Wallach. 2020. Power comes from language (technology): An analysis of "bias" in NLP. In ACL Proceedings.

[20] Abderrahim Beni-Hssane, Mohammed Kasri, and Marouane Birjali. An in-depth review of sentiment analysis trends, problems, and approaches. Received July 1, 2020; revised March 25, 2021; accepted May
10, 2021; accessible online May 14, 2021; record
version May 18, 2021.

[21] Sentiment analysis performance evaluation and comparison using deep learning techniques A. Pasumpon Pandian is the dean of research and development at CARE College of Engineering in Trichy, India. Online ISSN: 2582-2640 Sent on May 17, 2021; edited on June 7, 2021 Accepted on June 26,
2021; published on July 3, 2021.

[22] Olga, Radiuk, Pavlo, Pavlova, Hrypynska, and Nadiia.A group machine learning method for sentiment analysis in Witter. Published on July 17, 2022.

[23] Sentiment analysis in action. Lei Lei and Dinlin Liu, University Press, Cambridge, 2021.

[24] Sergio Consoli, Sebastiano Manzan, Elisa Tosetti, Luca Barbaglia, and Luca Tiozzo Pezzoli. Sentiment Analysis of Economic Texts Using Lexicons, 23 Pages Dates: May 11, 2022, May 13, 2022, and May 13,
2022.

[25] Patil, Ratna, and Sharavari Tamane. "A Comparative Study on the Assessment of Classification Algorithms for Diabetes Prediction." International Journal of Electrical and Computer Engineering (IJECE), vol. 8, no. 5, October 1, 2018, p. 3966.