UPES - University of Petroleum and Energy Studies

Department of Computer Science and Engineering

Case Study on Natural Language Processing in Data Science

Submitted by:

Neha Kumari
Sap ID:500092077

Professor & Faculty In-Charge:

Dr. Gourav Bathla

ABSTRACT

In the present state of digital world, computer machine do not understand the human's ordinary language. This is the great barrier between humans and digital systems. Hence, we found an advanced technology that provides information to the users from the digital machine. However, natural language processing (i.e. NLP) is a branch of AI that has significant implication on the ways that computer machine and humans can interact. NLP has become an essential technology in bridging the communication gap between humans and digital data. Thus, this study provides the necessity of the NLP in the current computing world along with different approaches and their applications.

1. INTRODUCTION

Data is increasing at an alarming rate. A large portion of the data available today is in the form of text. Natural Language Processing is a popular branch of AI which helps Data Science in extracting insights from the textual data. Following this, Industry experts have predicted that there will be a huge demand for Natural Language Processing professionals in the near future.[1]

M. Maxson, said that, in future most of the useful information will be in unstructured form. The future BigData will be the combination of both structured and unstructured data and utilizing inherent data patterns that integrate from data itself and not from police imposed on data-sets by humans. It has been frequently noted that NLP predominantly is utilized to analyse, retrieve and summarize the pertinent data from large sets of data available. An exploration of NLP concept was introduced in 1950 when Turing-test on computer machine and intelligence was introduced [2]. It was able to exhibit intelligent behaviour similar to, or non-differentiable from, that of a person. NLP need a combination of verbal and computational knowledge. This can be done for several languages for e.g. English; several challenges incurred when the information extraction contain paragraph phrasing, metaphors, idioms and rhetoric etc. [3].

Language disambiguation is the art to parse and extract the information in NLP. Analyse sentence-parsing, words identification using lexicon, Named entity recognition (NER) so on and extraction are the series of operations to be performed in NLP. Different approaches [4] example: NER, wrappers in web-corpus, tagging of parts of speech has been explored. The intention of information extraction is to extract the information for pre-determined blocks for a specific frame [5]. Generic models for NLP exploiting word joint probability and its tag or label, enhancing to Hidden-Markov approach and trigram markov approach has been introduced in [6]. It also defines usage of pseudo random words and their functioning for NLP of unidentified words for small dataset. In [7] Lapata et.al have illustrated about web-based approach for NLP that provided the knowledge of isolation of compound nouns, disorderly arrangement of adjectives, web interpolation and mass counts to maximize the flexibility. Thus, NLP falls into the domain of AI with the aim of understanding and making meaningful information in the human language [8].

Khalid et al. [9], presented identifying the multi-word Urdu from a number of corpora. The improvement of this scheme C-value method and NLP is utilized along with Urdu WordNet. The outcomes of this method comprises of 735 synsets containing only multi-word- Urdu terms from Urdu-WordNet. Jimenez et al. [10] have proposed to predict the data or searching information using Big Data predictive analytics in social media. The result discovered and analyzed present status of the research that has been urbanized so far by educational background. Elahinia et al. [11], evaluated a prediction of complex and multidimensional health care problem using NLP technique.

Slater et al. [12], developed an effective measure of student appointment in tuition or class. The technique NLP utilized to identify the different linguistic features such as mathematics problems on students affective states during work in a mathematic tutor. The outcome is measured with a linguistic feature, and student problems on online tutor are evaluated. Calapodescu et al. [13] presented a semi-automatic de-identification of hospital

discharge summaries. The proposed system is achieved with significant improvements for the annotation of the documents in quantitative, qualitative and homogeneous results. The work of Broniecki et al. [14] have focused on the upcoming arriving form of data and to make international practitioners grow. The presenting new data work on large question text and also worked on specific approaches. The outcomes are stable on human derived coding and complete on correlate/predict and achieve the data program.

In the study of Zitnik et al. [15] have concentrated on the various types of applications and tools that is required for machine learning in Natural language process. A new arrived application called toolkit provide an end to end text investigation by the use of JavaScript technique based function. The state of the arts are comes in the terms of the result. The work of Baby et al. [16] have focused on hose control application in which the user can interface with web application or a newly arrived chatbot tools its control on all house electrical equipment. The presenting chatbot technique follow the statement of text information that's text by user in the form of command and then control the all-electric equipment at house by using the motion sensors.

The study of Kaushik et al. [17] have focused on issues of NLP in which they discussed on deep learning which solves the problem of N-language processing. By the effective deep learning the secret issues like output vector and input word are removed. By the adding of Cross Language future work is done.The study of Ahmed et al. [18] have concentrated on poetry analyzer on N- language processing which used different types of language and its consider the computer based approach. This method considers all types of literary tested before presenting the final result is obtained.The work of pole et al. [19] have focused on the new proposed approach for the instruction of grouping documents in which search results are in the terms of better way-group fuzzy apparatus for accurate compilation by the using of weighted method. In the study of Bartero et al. [20] described anomaly identification problem. To resolve this issue introduced a Word embedding method based on Google's Word2vec algorithm. The results illustrated 90% accuracy and efficient computation.

Suhaimin et al. [21] described a challenge which occurs in sarcasm detection into the texts. The outcomes show that a group of prosodic, syntactic and pragmatic structures developed an effective performance with F-measure result of 0.852. The study carried out by Alami et al. [22] has presented a novel Semi-automated methodology to develop a sequential structure from user necessities written in Arabic. The result shows that proposed approach provides an effective Unified Modelling Language (UML) diagrams from Arabic user requirements. Zhuang et al. [23] presented a Stroke Level approach to influence Chinese stroke for learning the Chinese word representation. The experimental outcomes show that presented method is superior for Chinese Language processing. To reduce this type of problem the Deshmukh et al. [24] has developed Voice-based application to the users. The outcomes display the proposed system that provides correct answer to the user. It offers fast processing and maximum response capability to the users.

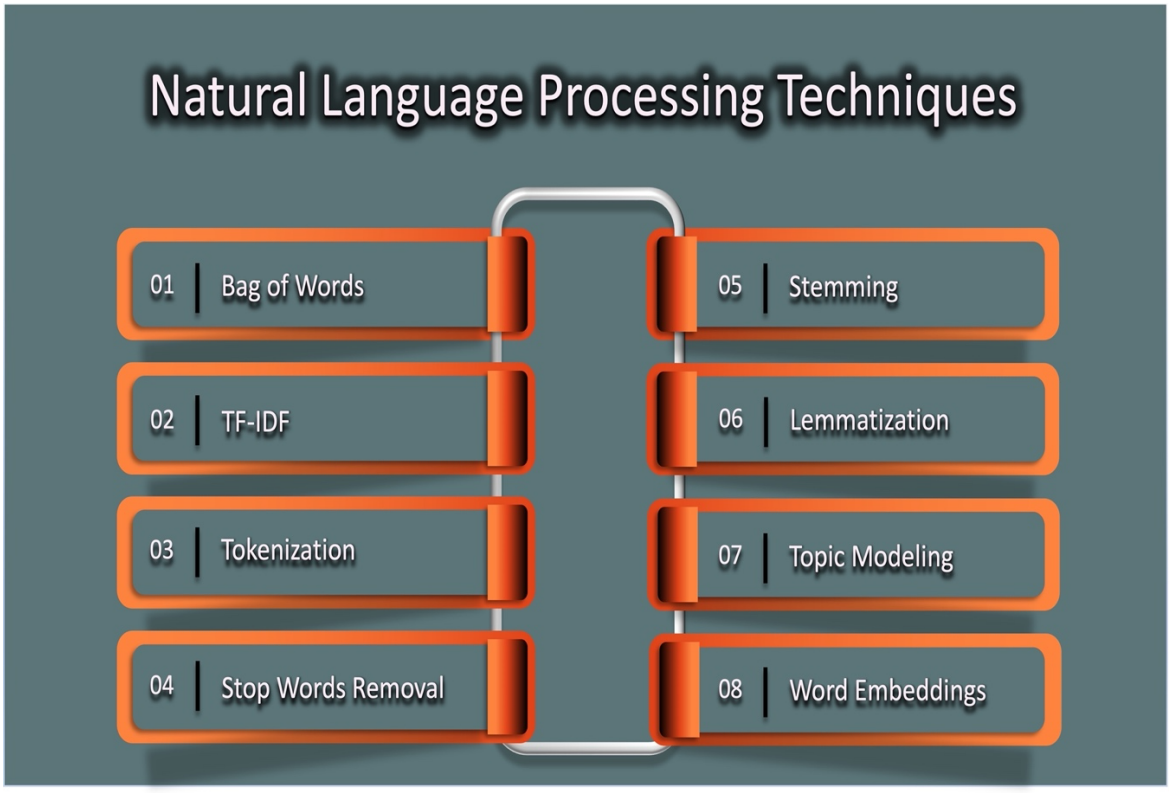
2. NATURAL LANGUAGE PROCESSING

2.1 What is NLP

Natural Language Processing or NLP is a branch that focuses on teaching computers how to read and interpret the text in the same way as humans do. It is a field that is developing methodologies for filling the gap between Data Science and human languages.
Everything we speak or express holds great information and can be useful in making valuable decisions. But extracting this information is not that easy as humans can use a number of languages, words, tones, etc. All these data that we are generating through our conversations, tweets, etc is highly unstructured. The traditional techniques are not capable of extracting insights from this data. But thanks to the advanced technologies like machine learning and NLP that have brought a revolution in the field of Data Science. [25]

Many areas like Healthcare, Finance, Media, Human Resources, etc are using NLP for utilizing the data available in the form of text and speech. Many text and speech recognition applications are built using NLP. For example, personal voice assistants like Siri, Cortana, Alexa, etc.

2.2 NLP techniques in Data Science



Some of the most widely used NLP techniques in Data Science :

(i) Bag of Words

This model counts the number of words in a piece of text. This model works by generating an occurrence matrix for the sentences. The underlying grammar and the order of words are not considered while generating the matrix.
These occurrences or counts are then fed into a classifier as features. A compound sentence is a sentence formed with two or main clauses joined by a coordinating conjunction. A simple sentence is a sentence formed with one main clause. Now let's generate the occurrence matrix for this:

	Two	Compound	Simple	Sentence	More	Formed	One	Clauses	…
A Compound sentence is a sentence with two or more main clauses joined by a coordinating conjunction.	1	1	0	2	1	1	0	1	…
A simple Sentence is a sentence formed with one main clause.	0	0	1	2	0	1	1	1	…

This approach is very simple to understand but it has several drawbacks also. This model gives no idea about the semantics and the context in which the words are used. Also, some words like "a" or "the" which appear frequently but are not that important may create noise during analysis. Another problem is that in the above example, the word "then" holds more weight than the word "universe" i.e words are not weighted according to their importance.

To overcome these issues, the use an approach called Term Frequency-Inverse Document Frequency (TF-IDF).

(ii) Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency or TF-IDF overcomes the drawbacks of Bag of Words by using a weighting factor. It uses statistics for calculating the importance of a word in a document. Let us understand the statistics of TF-IDF.

TF or Term frequency: It measures the frequency of a word in a document. This is calculated by counting the total number of occurrences of the word and dividing it by the total length of the document.
IDF or Inverse Document Frequency: It measures the importance of a word in a document. For example, words such as is, a, of, etc which occur frequently in the document but they do not hold much importance as they are not adjectives or verbs. Thus this technique assigns a weight to any string according to its importance. It is calculated by taking the log of the total number of documents in the dataset divided by the number of documents containing that particular word (also 1 is added to the denominator so that it is never 0).

TF-IDF: Finally it calculates the importance of any word by multiplying the TF and IDF terms i.e TF*IDF.
Thus the words having more importance are assigned higher weights by using these statistics. This technique is mostly used by search engines for scoring and ranking the relevance of any document according to the given input keywords.

(iii) Tokenization in NLP

This is one of the NLP techniques that segments the entire text into sentences and words. In other words, we can say that it is a process of dividing the text into segments called tokens. This process discards certain characters like punctuations, hyphens, etc. The main purpose of tokenization is to convert the text into a format that is more convenient for analysis. Let us apply this to an example. The result will be like:

The	night	is	black	and	the	forest	has	no	end
a	million	people	tread	it	in	a	million	ways	

In this case, it was quite simple as we have split into blank spaces, but this will not be the case always. Sometimes splitting into blank spaces may break such words into different tokens that should be considered as one token (for example New York or New Delhi).

My	Name	Is	Neha	I	Live	New	Delhi
My	Cat	Loves	To	Drink	Milk	All	Day

Another problem with tokenization is the removal of punctuations. Sometimes it may lead to complications. For example, in Mr., the period following the abbreviation should be a part of the same token and should not be removed. Because of this, a large number of problems arise while applying tokenization to biomedical text domains having a number of hyphens, parentheses, and punctuations.

(iv) Stop Words Removal

The objective of Stop Words Removal is similar to tokenization. In this process, the common words which occur most frequently but adds a very little or no value to the result are removed from the text. For example, the common prepositions like and, the, a, etc of the English

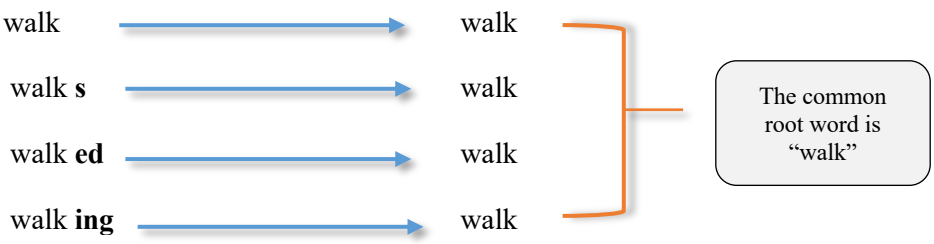
language. The main purpose behind this is to minimize the noise so that we can focus on the words holding important meaning during the analysis.

Stop words removal can be easily carried out by looking up in a predefined list and removing those words. This helps to free up space and improve performance and processing time. But it should be kept in mind that there is no universal list of the stop words. You can build or generate it using scratch. Another effective approach is to take all the predefined stop words in the beginning and then add new words according to your requirements. Sometimes some important information is lost in this method.

For example, it would be problematic if we discarded words like “not” while performing sentiment analysis. In such cases, you should prefer a list having minimal stop words at first and then add more words later on.

(v) Stemming in NLP

The main objective of this NLP techniques is to reduce the words to their root form. This technique works on the principle that certain words having slightly different spellings but nearly the same meaning should be in the same token. Thus, we chop off the affixes of the words for efficient processing. For better understanding, consider the following example.



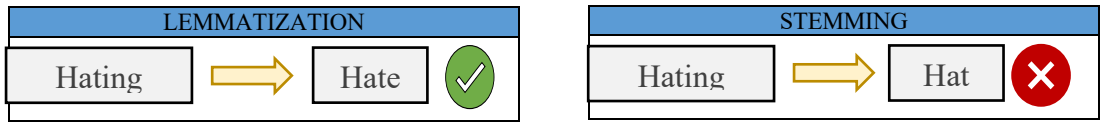
In the above example, all forms refer to the same word “walk”. So it would be better if we map these words to the same token during analysis. But how to decide whether to chop the affix or not. Consider the following example, For dealing with such issues, you can refer to some lists containing common affixes and perform stemming accordingly. There are a number of limitations in stemming but on the other hand, it is very efficient in terms of speed and performance.

(vi) Lemmatization in NLP

Lemmatization has an aim similar to stemming. This NLP technique aims at reducing the inflected forms of a word to their root form and group them together. For example,
a) came is converted to come (past tense changed to present tense)
b) worst is changed to bad (synonyms to their simpler form)
The aim of stemming and lemmatization is quite similar but the approaches are different.
In Lemmatization, we convert the words into lemma which is the dictionary form of the word.

For example, “Swims”, “swimming” are forms of the word “swim”. So “swim” will be the lemma for these words.

The following image properly indicates the difference between Stemming and Lemmatization.



Lemmatization has the capability to differentiate between words based on their context. For example words like, a bank which refers both to a financial institution and to land alongside a river body. As we have seen that Lemmatization performs better than Stemming but it also requires more computational power than Stemming.

(vii) Topic Modeling in NLP

Topic Modeling is a technique in NLP that extracts main topics from the text or document. It works on the assumption that each document is a group of topics and each topic is a group of words. We can relate it with the Dimensionality Reduction. Because in this technique also, we reduce the large text into a smaller number of topics.

The technique of Topic Modeling employs in various Data Science applications like Data Analysis, Classification, Recommender systems, etc. One of the popular techniques for Topic Modeling is LDA which stands for Latent Dirichlet Allocation. LDA is an unsupervised technique for finding the appropriate collection of topics in a document. LDA works in the following way:

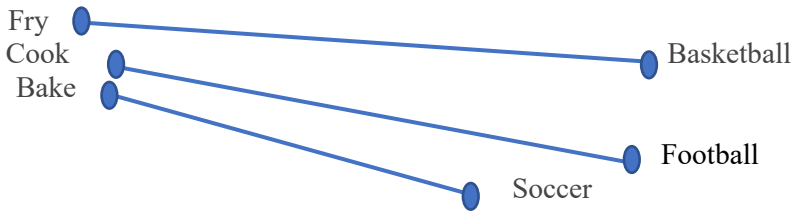
Firstly the user defines the number of topics a document should have. The algorithm will then divide the document into topics in such a way that the topics should include all the words in the document.

The algorithm then iteratively assigns the words to any topic based on its probability of belonging to that topic and the probability that it can regenerate the document from those topics.

(viii) Word Embeddings in NLP

It is a technique for representing words of a document in the form of numbers. Its representation should be such that similar words have a similar representation. The modern approaches represent the words as real-valued vectors.

The length of all word vectors is the same but each vector has a different value. The distance between vectors indicates the similarity between them.



In the above example, the distance between the words fry and cook is less as they are similar while the distance between cook and football is more as they are not similar.

2.3 Real Life NLP Application

There are some examples which shows how businesses are integrating NLP with Data Science for better results:

- In 2015, Uber launched its messenger bot on Facebook Messenger. The aim was to reach more and more customers for collecting more data and Facebook was the best possible way to connect people through social media. This bot helped them in providing a better and personalized customer experience by analyzing the customer data. This bot provided the users with easy and quick access to the service which eventually helped them in gaining more users. The bot also helped them to generate more revenue through advertisements.
- Many e-commerce businesses are using Klevu, a smart search provider based on NLP for providing a better customer experience. This smart search provider automatically learns from the user interactions in the store. It performs various functions like search autocomplete, the addition of relevant contextual synonyms, etc. It also uses the insights gained from the textual data for providing personalized search recommendations.
- In 2016, Mastercard also launched its Chatbot on Facebook Messenger. The aim of this chatbot was to provide various customer support services like an overview of their spending habits, available benefits, etc by analyzing their data. This helped them to provide a better customer experience. This initiative of chatbot also saved their expenses of developing a separate app for customer support.

3. KLEVU SMART SEARCH

3.1 Introduction

Klevu is the leader in Search and Category Navigation, Cofounded in Finland in 2013 by Nilay Oza, Niraj Aswani, and Jyrki Kontio, Klevu aims to connect people to products they wish to buy. Through AI and natural language processing-powered technologies, the company's products help merchants deliver relevant experiences powered by behaviour[26].
It is easy to configure, optimize and maintain and fully integrates with major e-commerce platforms in just hours.

Klevu Smart Search is a powerful and extensible e-commerce search solution that delivers search results based on shopper intentions and behaviour, in real-time. Klevu is a self-learning technology that boasts world class AI encompassing leading edge Machine Learning (ML) and Natural Language Processing (NLP).[27]

Version	Release Date	
1.0.0	Aug 12, 2020	Repository size is 6.84 KB
1.0.1	Nov 4, 2021	
1.1.0	Mar 30, 2022	

Most of the features provided by Klevu Search are language independent. This includes features such as its capability to self-learn the ranking of products based on consumer activity, fuzzy match to handle error tolerance, synonyms, URL redirects, product promotions, normalisation of model numbers and SKU codes, accented characters, handling stopwords, partial match etc.

There are a few features that are language dependent.

(i) Normalising inflected words: This feature allows us to normalise inflected words. For example, plurals to singulars (e.g. batteries -> battery), identifying base forms (e.g. singing -> sing). Indexing both the inflected form and the normalised form allows shoppers a flexibility to search for any form. Currently, this feature is supported for the following languages: English, Finnish, Swedish, Danish, Hungarian, Russian, Turkish, Dutch, French, Italian, Portuguese, Spanish, German, Norwegian, Romanian.

(ii) Automated Enrichment: Based on the context of a word, our automated enrichment algorithm identifies relevant synonyms, which it then adds to the search index. For example, if a word is “smart”, we may add additional words such as “intelligent” and “clever” to the search index. If the word is a “curtain”, we may add the word “drape”. Automated addition of relevant synonyms allows shoppers to locate a product not only by using the original word but with additional synonyms added by our algorithm. Currently, this feature is supported for the following languages: English, Finnish, Swedish, Spanish, and French languages.

(iii) Decoupling Compound Words: If a word is “rottvinglas” (i.e. red wine glass in English), we are able to decouple it and index it as three separate words “rott”, “vin” and “glas”. This allows shoppers to fire a query such as “glas for rott vin” (i.e. glass for red wine). Currently, this feature is supported for the following languages: English, Finnish, Swedish, Estonian, German, Spanish and Danish.

(iv) Advanced Query Processing: Queries such as “formal tops under 100 USD” requires understanding the intent behind the query and fetch the appropriate products. Currently, the

feature is available for the following languages: English, Finnish, Spanish, French, and Swedish languages.

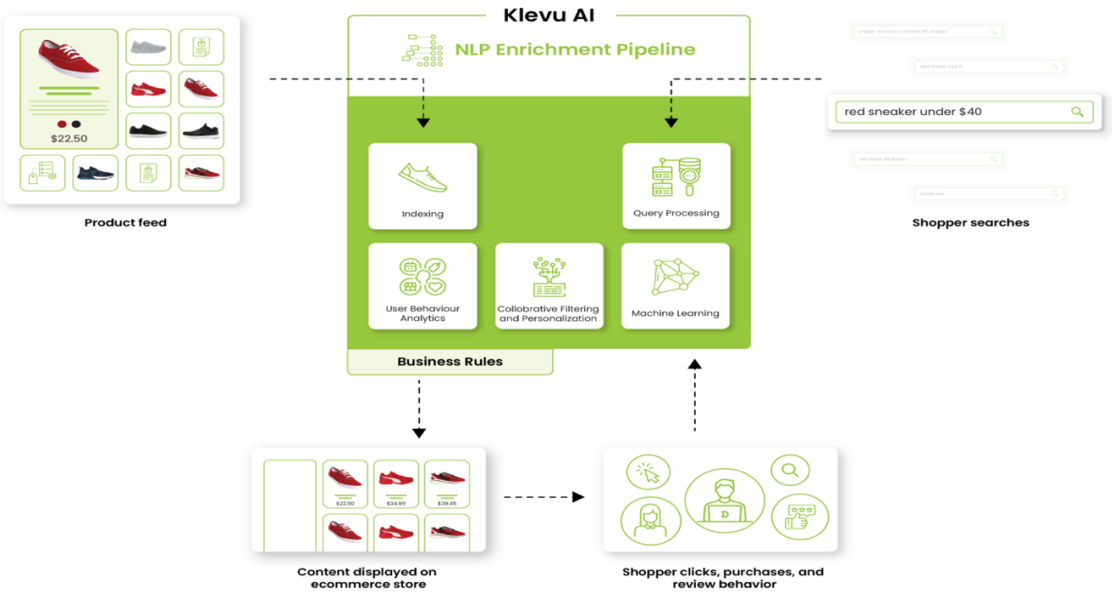
3.2. Klevu Search Features

Following are the features of klevu smart search

- Self-learning search
- Dynamic filters
- Search results page
- Fully customizable
- Product promotions
- 100% cloud hosted, CDN backed and secure backups
- Error tolerance: Automatically take care of spelling mistakes
- Identifies stop words, inflections and more
- Populates trending and popular searches
- Synonym dictionary
- URL redirect
- Comprehensive data tracking & analysis
- Easy integration process

3.3 Working of Klevu AI

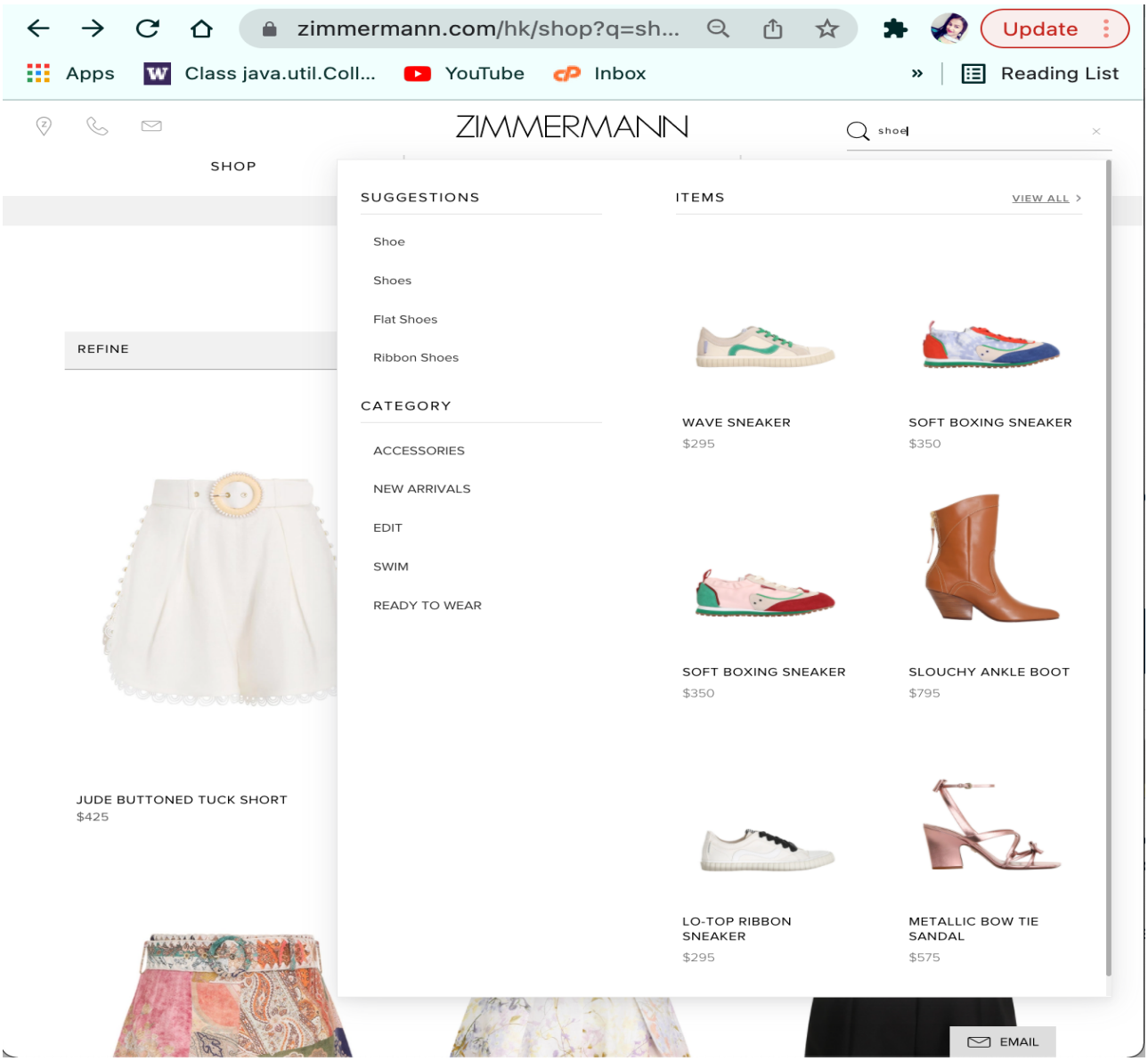
Klevu AI encompasses linguistics and continuous machine learning to improve online shopping experiences. Klevu shines in a crowded product discovery market because of the comprehensiveness of the AI offering, and how it impacts user experience.[28]



NLP: Semantic Query Processing

Klevu processes understands what the shopper means when they search, considering the context of the store, to improve the quality of the results.

Natural language processing (NLP) is one of the core components of Klevu, helping to extract more meaning and context from search queries and match terms to results that other technologies are unable to match.
You can see an example of how NLP can support search on the Zimmermann store (a Klevu client) below.



As you can see, the product names doesn't contain the word "shoe" but the system is able to understand which products are shoes. So it is using synonyms of search word by importing wordnet. WordNet is a lexical database for the English language, which was created by Princeton, and is part of the NLTK corpus.[29]

Klevu also uses machine learning to promote products based on how users are interacting with them, providing more real-time promotion of items based on their popularity and effectiveness. The key interactions that Klevu uses are completed purchases, add-to-carts, and clicks. These actions give us an understanding of popularity. This gives us a picture of short-term and long-term popularity and allows us to add a layer of rankings on top of just boosting specific categories or items.

3.4 How Klevu smart search is different from other traditional Search engine?

In the past, keyword-based search algorithms have often led to something of a 'chicken and egg' situation, with website visitors trying to 'guess' what terms they need to enter in the search box, in order to get the results they want from their search. Clearly, this is not an ideal approach, and, historically, it has led to high levels of frustration for online shoppers.

With NLP-based search, however, the focus switches from keywords to actual meaning. Taking a semantic approach means that search results have a 'connection' to the search terms, rather than having to actually contain those search terms. With over 1 million words in use in the English language, it's not hard to see why a more context-driven search algorithm is more useful when it comes to retrieving accurate results.[30]

There are two things that makes Klevu smart search different from other traditional search engine (i) The Natural language Processing, and (ii) Self-learning Capabilities

It uses NLP, to go a lot deeper in matching products to query, which help to find more accurate search results

Other key aspects that differentiate:

- Search merchandising capabilities (boosting specific products and categories, etc.)
- The integration (It has plug and play integrations with Magento 1, Magento 2, and Shopify and we can integrate with any other platform relatively quickly and easily as well)
- Its support (this is something they are very proud of)

3.5 E-commerce website which is using Klevu Smart Search

- Magento
- Yamaha
- Made.com
- Puma
- Spar
- Cox & cox
- CASE.MATE
- Avon

3.6 Benefit of using Klevu Smart Search

- Puma increases search-led conversion by 52% with Klevu AI search and merchandising
- 80% increase in direct-to-consumer sales, and 10% e-commerce conversion rate

CONCLUSION

From this study, we have seen the various popular NLP techniques used in Data Science. We have also seen how different companies are using Data Science and NLP for improving their business. These NLP Techniques are very helpful while using NLP in data science.
Here I have discussed about Klevu smart search based on Natural Language Processing(NLP) technique which also uses self-learning. It works best for e-commerce because it learns by watching how shoppers interact with search on the store.
In addition to providing the basic auto-complete search function, Klevu automatically adds contextually relevant synonyms to a catalogue that can result in 3x the depth of search results. The software also provides personalized search, offering products that customers previously interacted with or products that are trending.
With NLP-based search, however, the focus switches from keywords to actual meaning. Taking a semantic approach means that search results have a ‘connection’ to the search terms, rather than having to actually contain those search terms. With over 1 million words in use in the English language, it’s not hard to see why a more context-driven search algorithm is more useful when it comes to retrieving accurate results.

BIBLIOGRAPHY

[1] https://techvidvan.com/tutorials/nlp-techniques-in-data-science/

[2] "Natural Language Processing," [Online]. Available: https://en.wikipedia.org/wiki/Natural_language_processing.

[3] "Syntax- English sentence structure," [Online]. Available: esl.fis.edu/learners/advice/syntax.htm.

[4] Ratinov, Lev, and Dan Roth, "Design challenges and misconceptions in named entity recognition," In Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009), pp. 147-155, 2009.

[5] R. J. Mooney and R. Bunescu, "Mining Knowledge from Text Using Information Extraction," SIGKDD Explorations, vol.7, no. 1, pp. 3-10, 2005.

[6] M. Collins, "Chapter 2: Tagging Problems, and Hidden Markov Models," Course note for NLP, Columbia University, 2013.

[7] M. Lapata and F. Keller, "Web-based models for Natural Language Processing," ACM Transactions on speech and Language Processing, vol. 2, no. 1, pp. 1-30, Feb. 2005.

[8] N. Ranjan, et al., "A Survey on Techniques in NLP," International Journal of Computer Applications, vol. 134, no. 8, pp. 6-9, Jan. 2016.

[9] K. Khalid, et al., "Extension of Semantic Based Urdu Linguistic Resources Using Natural Language Processing," 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th International Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Orlando, FL, pp. 1322-1325, 2017.

[10] J. L. J. Marquez, et al., "Challenges And Opportunities In Analytic-Predictive Environments Of Big Data And Natural Language Processing For Social Network Rating Systems," in IEEE Latin America Transactions, vol. 16, no. 2, pp. 592-597, Feb. 2018.

[11] H. Elahinia, et al., "Predicting medical nonadherence using natural language processing," 2017 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, pp. 1-4, 2017.

[12] S. Slater, et al., "Using natural language processing tools to develop complex models of student engagement," 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, pp. 542-547, 2017.

[13] I. Calapodescu, et al., "Semi-Automatic De-identification of Hospital Discharge Summaries with Natural Language Processing: A Case-Study of Performance and Real-World Usability," 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Exeter, pp. 1106-1111, 2017.

[14] P. Broniecki and A. Hanchar, "Data Innovation for International Development: An overview of natural language processing for qualitative data analysis," 2017 International Conference on the Frontiers and Advances in Data Science (FADS), Xi'an, pp. 92-97, 2017.

[15] S. Žitnik, et al., "nutIE—A modern open source natural language processing toolkit," 2017 25thTelecommunication Forum (TELFOR), pp. 1-4, 2017.

[16] C. J. Baby, et al., "Home automation using IoT and a chatbot using natural language processing," 2017 Innovations inPower and Advanced Computing Technologies (i-PACT), pp. 1-6, 2017.
[17] A. R. Sharma and P. Kaushik, "Literature survey of statistical, deep and reinforcement learning in natural language processing," 2017 International Conference on Computing, Communication and Automation (ICCCA), pp. 350-354, 2017.
[18] M. A. Ahmed and S. Trausan-Matu,"Using natural language processing for analyzing Arabic poetry rhythm," 2017 16th RoEduNet Conference Networking in Education and Research (RoEduNet), pp. 1-5, 2017.
[19] K. R. Pole and V. R. Mote, "Improvised fuzzy clustering using name entity recognition and natural language processing," 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), pp. 123-126, 2017.
[20] C. Bertero, et al., "Experience Report: Log Mining Using Natural Language Processing and Application to Anomaly Detection,"2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE), Toulouse, pp. 351-360, 2017.
[21] M. S. M. Suhaimin, et al., "Natural language processing based features for sarcasm detection: An investigation using bilingual social media texts,"2017 8th International Conference on Information Technology (ICIT), Amman, pp. 703-709, 2017.
[22] N. Alami, et al., "A semi-automated approach for generating sequence diagrams from Arabic user requirements using a natural language processing tool,"2017 8th International Conference on Information Technology (ICIT), Amman, pp. 309-314, 2017.
[23] H. Zhuang, et al., "Natural Language Processing Service Based on Stroke-Level Convolutional Networks for Chinese Text Classification,"2017 IEEE International Conference on Web Services (ICWS), Honolulu, HI, pp. 404-411, 2017.
[24] S.Deshmukh, et al., "Sia: An interactive medical assistant using natural language processing,"2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), Jalgaon, pp. 584-586, 2016.
[25] https://techvidvan.com/tutorials/nlp-techniques-in-data-science/
[26] https://venturebeat.com/2021/04/08/klevu-raises-12m-for-ai-that-personalizes-ecommerce-search/
[27] https://libraries.io/packagist/klevu%2Fmodule-mysqlcompat
[28] https://www.klevu.com/capabilities/ai-and-nlp/
[29] https://www.marketingaiinstitute.com/blog/klevu-uses-artificial-intelligence-to-help-ecommerce-stores-sell-more
[30] https://info.klevu.com/blog/benefits-of-using-natural-language-processing-in-ecommerce-site-search