VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Natural Language Processing

CSE4022 (EPJ)

Fall Sem 2021

Slot: B2

Faculty: Prof. Rajeshkannan R

SCOPE

Analysis of social media to provide actionable intelligence for automobile brands

TEAM MEMBERS	REGISTRATION NUMBER
Yuvraj Singh Pathania	19BCE0561
Nikhil Kumar Parashar	19BCE0076
Dussa Lalith Kumar	19BCE0532

Abstract

Natural language processing (NLP) is a branch of linguistics, computer science, and artificial intelligence concerned with computer-human interaction, particularly how to design computers to process and evaluate massive volumes of natural language data. The goal is to create an algorithm that can "understand" the contents of papers, including the nuances of language in context. The system can then extract accurate information and insights from the papers, as well as categorize and organize them. Some of the applications of NLP involve - Text and speech processing, Morphological analysis, Syntactic analysis, Lexical semantics (of individual words in context), Relational semantics (semantics of individual sentences), Discourse semantics (beyond individual sentences), etc.

AI(NLP) is already able to make a significant impact on the automotive industry value chain by affecting every part of the business—and its impact is only set to increase. AI will enhance quality, increase efficiency, decrease risk, and, driven by data, AI will improve the type and level of service that automakers can deliver to their existing and prospective customers, ensuring that everyone benefits from the best possible product.

The project we are building takes advantage of an unlimited resource,i.e social media to extract information and produce commercially beneficial analysis.
We intend to use multiple NLP techniques to build a hybrid analysis model, which will specifically help the automobile companies to extract valuable information from social media posts, regarding their product and other similar products in the market which is in direct and indirect competition to theirs.

We aim to present a comprehensive dashboard with detailed analysis, using which companies can get an excellent overview from a high-level perspective as well as the little details that stand out, they can understand how well their product is performing, and make intelligent business and technology decisions while releasing new products.
The relative analysis that we will provide will help them gain an edge over their competitors, in terms of technology that is most important to people and help them get an overall sentiment of the consumer, and make appropriate business decisions to win in the ultimate race of selling their products to more customers and also help them retain customers longer.

Introduction

The evolving market landscape:

A product launch is a moment of pride for any automobile brand, it is essential as the brands operate in a hyper-competitive environment, where the customer's loyalty is very fickle. The conventional forms of marketing usually involve purchasing in print and electronic media, but the launch phase is very critical as it sets the course for the product being launched and can make or break its sales/hype cycle.

In the current scenario of marketing considering the internet and social media, the landscape of advertising and reaching the customer has changed. The customers are now closer to the marketer than ever before. This has therefore massively amplified and magnified the inherent product or marketing characteristics, but only as interpreted by the masses.

This brings up questions such as, how does a marketer know how his message is being interpreted by the consumers? How do they make sure that it is developing a favorable relationship and that it would be the next thing to go viral or take the market by surprise, how to make sure the advertising is reaching the right target audience and how to prevent people from brand switching.

Marketing to the Opinionated buyer:

According to a report compiled by the Content Marketing Institute of North America and the Association for Data-Driven Marketing, 96 percent of B2C companies used social media for focused marketing in 2018.(According to the Content Marketing Institute, 2018).It is also well known that numerous companies use social media as part of their customer service, activities aimed at resolving user questions and complaints as quickly as feasible. Customers prefer to churn, these days, rather than downloading specialized apps or using company-specific tools to contact customer support(2019, Appel).

These social media companies use data analytics and offline engagement to better understand how consumers react to content marketing and create feedback loops to provide the best outcomes for their clients.

The success of marketing initiatives is critical to the businesses that launch them. Brand and campaign awareness are frequently used to determine the effectiveness of ad campaigns, with metrics such as increased followers and mentions following the campaign, subjectivity and sentiment scores of responses, retweets and tags, view rate and view time, as well as overall brand sentiment, which refers to the general perception of the brand on social media, not to mention the actual sales figures of the companies in question being used (Pradipta Rani, 2011).

The role of Twitter in Content Marketing:

Twitter is a global messaging and social media network based in the United States that was created in March of 2006. Twitter is a microblogging service that allows users to share content in the form of 280-character messages known as 'tweets.' Because of its popularity among mobile device users, particularly smartphone users, Twitter has become omnipresent.
Twitter's uniqueness drew such a large audience that it now claims to be the most active social media outlet. In contrast, this audience's reliance on Twitter has grown to the point where 54% of users question the significance of an event if it is not trending on Twitter (Kantar Media, 2018).

Marketers have flocked to Twitter because of its alleged ability to provide 3x returns on marketing spends (AMIC, 2018). Its 335 million average monthly active users (AMIC, 2018) was startling, and when combined with a 53 percent higher likelihood of trying a new product than a non-user (AMIC, 2018), it was an opportunity too good to pass up for any company offering new items.

Information Extraction & Sentiment Analysis in Marketing:

Organizations are divided into two groups: trend setters or leaders who create innovative products and entice customers to buy them through effective communication and insight-driven processes, and second, organizations that are trend setters or leaders who create innovative products and entice customers to buy them through impactful communication and insight-driven processes. The second type of company is more reactive; they wait and watch, observing trends in consumer behavior, determining what works and what doesn't, and then launching a product or service in response to market demand. What these two have in common is that they both take risks in their businesses. It is possible to accept or reject their output. This is why they go to great lengths to validate their business concepts, goods, and conduct comprehensive tests in order to obtain community input.

Because the sampling or data available was unworkable, the basic premise of traditional methods of gathering customer sentiment was wasteful. With the inflow of data directly from the user's endpoint to the marketer, this is changing. Not only has sales data gone digital, but so has the voice of the client, as global internet access, smartphone usage, and e-commerce acceptance has increased tremendously. People currently express their delight, discontent, reviews, expectations, and motives for all of their purchases and plans more than ever before, to the point that it is nothing short of a treasure trove.

Information Extraction and sentiment analysis also assist brands in benchmarking their performance and increase competition among industry participants. Overall, data analytics aids firms in pursuing a more focused strategy for attracting and retaining customers.

Literature Survey

S.No	Paper/ Journal Title	Method/Algorithm	Challenges	Observations
1	Research on information retrieval model based on ontology, 2019	BM25, TF-IDF	Users face a significant challenge in obtaining correct and timely information when confronted with a large amount of data in the network.	In terms of iteration counts and average distance of the rank list, the genetic algorithm with simulated annealing approach is compared. The average distance is close to the overall ideal after 200 iterations.
2	Implementation of Text Based Information Retrieval Technique, 2020	Sequence Matcher	The data available on the internet in the form of blogs or text files is irrelevant, making information extraction difficult.	The paper only got the first twenty results from various precision queries, while the Bing API delivers over thirty-five.
3	Dependency graph for short text extraction and summarization, 2019	Knowledge Graph Belief Graph Dependency Tree	The majority of textual data is in the form of brief, fragmented texts that are difficult to visually extract due to the sparsity of the information, and the context is frequently unknown.	We demonstrated the usefulness of the belief graph, particularly in mining short text, by performance and practical studies in tests and utilizing real-world social media post datasets.
4	Sentiment analysis using product review data, 2015	SVM, Random Forest, Naive Bayes	Sentiment polarity categorization	Sentiment polarity categorization is a key challenge in sentiment analysis that is addressed in this study.
5	An analytical study of information extraction from unstructured and multidimensional big Data, 2019	Self-training with CNN-LSTM-CRF CNN+Bi-LSTM and CRF Grammar rules+MapReduce	With the increasing expansion of multifaceted, also known as multidimensional unstructured data, big data poses new issues for IE approaches. Traditional Internet Explorer systems are ineffective in dealing with this massive influx of unstructured big data.	With the tremendous growth of unstructured big data, it has been discovered that data analysis and mining are becoming increasingly challenging. Deep learning, with its generalizability, versatility, and potential to require less human participation, is a crucial player in this regard.

6	A Natural Language Processing and Deep Learning based Model for Automated Vehicle Diagnostics using Free-Text Customer Service Reports, 2021	BiLSTM, CNN	Initial issue identification and diagnostics are critical steps in increasing vehicle efficiency, safety, and stability. Several studies have looked into data-driven techniques to improve vehicle diagnostics utilizing available vehicle data in recent years.	The BiLSTM model had an initial accuracy of roughly 63%, which was slightly better than the other statistical models.
7	Research on Machine Learning Techniques for POS Tagging in NLP, 2019	POS Tagging, Naive Bayes, Max Entropy Classifier, BernoulliNB Classifier, Perceptron Tagger	This work aims to identify the primary kinds of activities that fall under Natural Language Processing and to comprehend the most often used machine learning approaches for these jobs.	Minor variations in content category, tag set selection, and training set size have been seen to have an impact on POS tagger accuracy.
8	Part of speech tagging: a systematic review of deep learning and machine learning approaches, Jan 2022	ML-based POS tagging DL-based POS tagging Hidden Markov Model	POS tagging still confronts challenges in boosting accuracy while reducing false-positive rates and tagging unknown terms, despite academics' best efforts. Furthermore, when tagging terms with distinct contextual meanings inside a sentence, the presence of ambiguity cannot be neglected.	To begin, the theoretical notion of NLP and POS tagging, as well as the numerous POS tagging methodologies, are thoroughly discussed using the examined research publications. The approach used by each article is discussed, as well as the strengths and weaknesses of each article in terms of the POS tagging model's capabilities and difficulty.
9	Part-of-Speech (POS) Tagging Using Deep Learning-Based Approaches on the Designed Khasi POS Corpus, 2021	BiLSTM BiLSTM with CRF Character-based embedding with BiLSTM	There hasn't been any kind of Khasi corpus established or formally developed until now.	As a result, the BiLSTM approach achieves 96.81 percent accuracy, 96.98 percent for BiLSTM with CRF methodology, and 95.86 percent for character-based with LSTM.
10	Identification of POS Tag for Khasi Language based on Hidden Markov Model POS Tagger, 2019	Hidden Markov Model	From the standpoint of NLP, Khasi lacks extensive research.	Aside from the correctly tagged terms, some inaccuracies were discovered in the experimental results.

Problem Statement

Problem Description:

To analyze and formulate multiple datasets from social media(Twitter), pertaining to each brand, using various data extraction and categorization techniques and perform **Information Extraction** analysis to categorize understandings and provide the automobile brands with a comprehensive dashboard, with actionable intelligence which can help them make decisions on multiple levels.

We intend to analyze the data extracted from social media(Twitter), in regard to automobile brands and consider products that fall in a certain monetary range; We would be using information retrieval, information extraction, named entity extraction, to build our dataset. (Along with some other datasets as mentioned below)
We would be using **POS Tagging, Pattern Matching, NER, Dependency graphs,** and other relevant techniques to formulate useful information. Additionally, we would **form opinion clouds,** and analyze the **keywords** generated, we then categorize data into various segments that pertain to the automobile industry, and then devise a comprehensive dashboard with detailed analysis to help the companies make informed decisions related to sales, operations as well as ideation.

Information Extraction

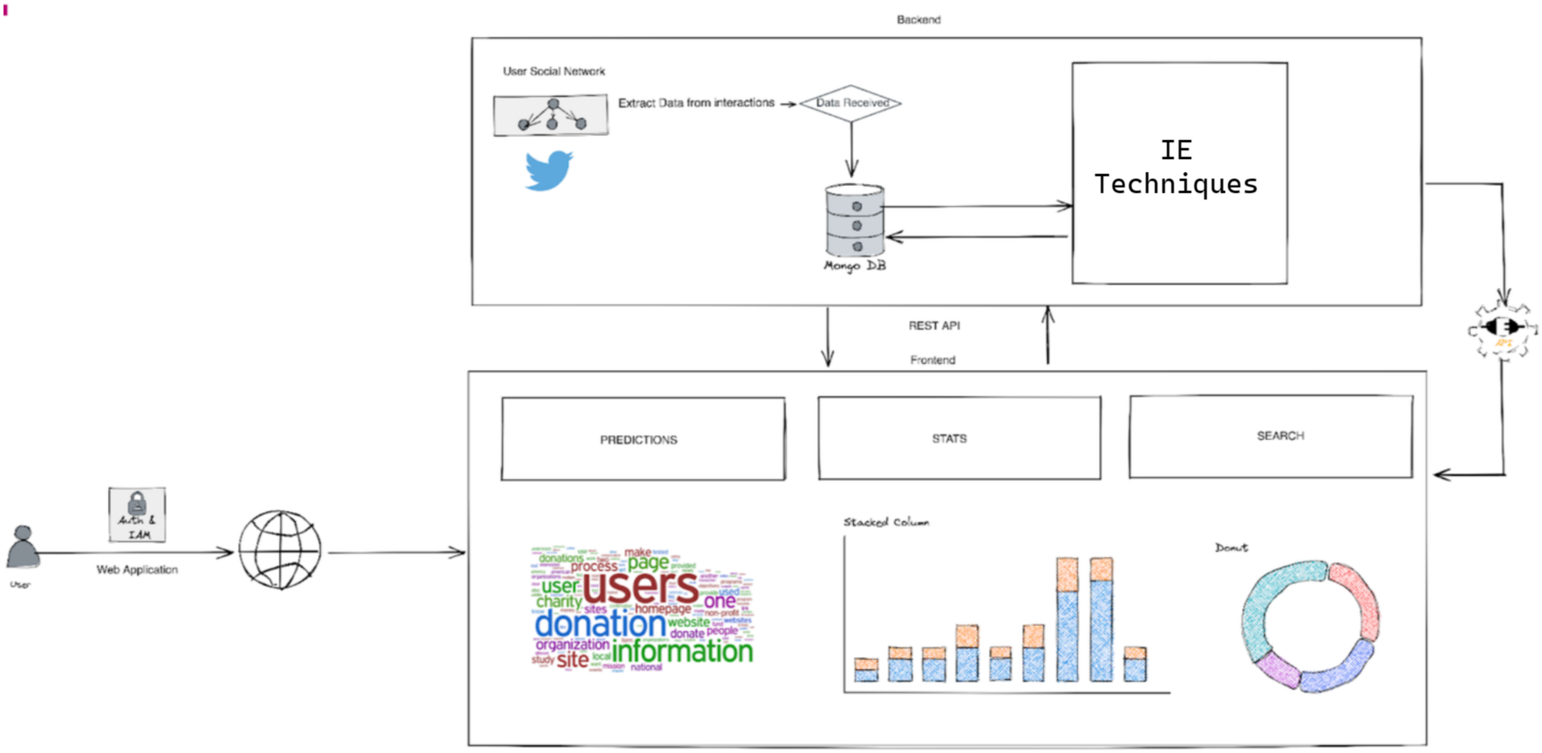
There is a lot of information in text data, but not all of it is relevant to us. We may be seeking entity names, while others may be looking for precise relationships between those entities. Depending on our needs, we have different intentions.

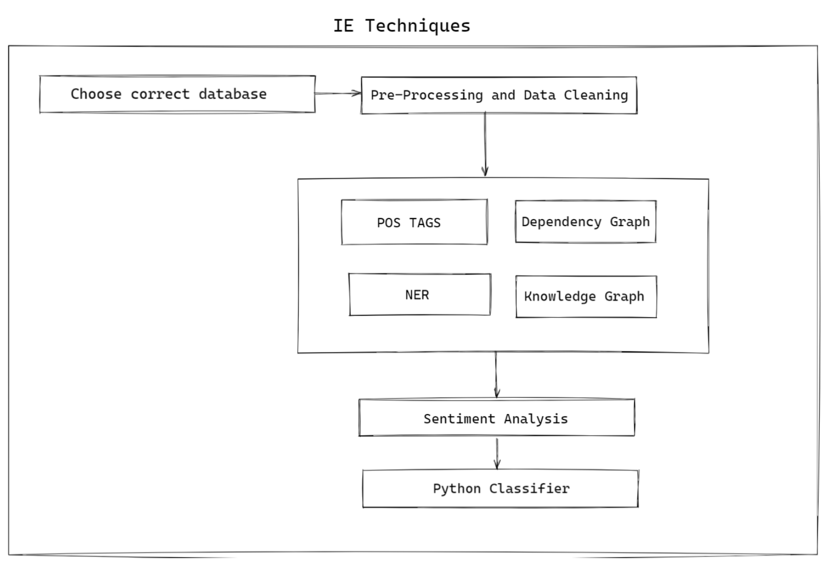
Simply put: The process of sifting through unstructured data and extracting vital information into more editable and structured data forms is known as information extraction.
.

As a result, devising an automated method of extracting information from textual material and presenting it in a structured fashion will enable us to gain several benefits while drastically reducing the amount of time we must spend skimming through text documents. This is precisely what information extraction aims to accomplish.

with information extraction NLP algorithms, we can automate the data extraction of all required information such as tables, company growth metrics, and other financial details from various kinds of documents (PDFs, Docs, Images, etc.).

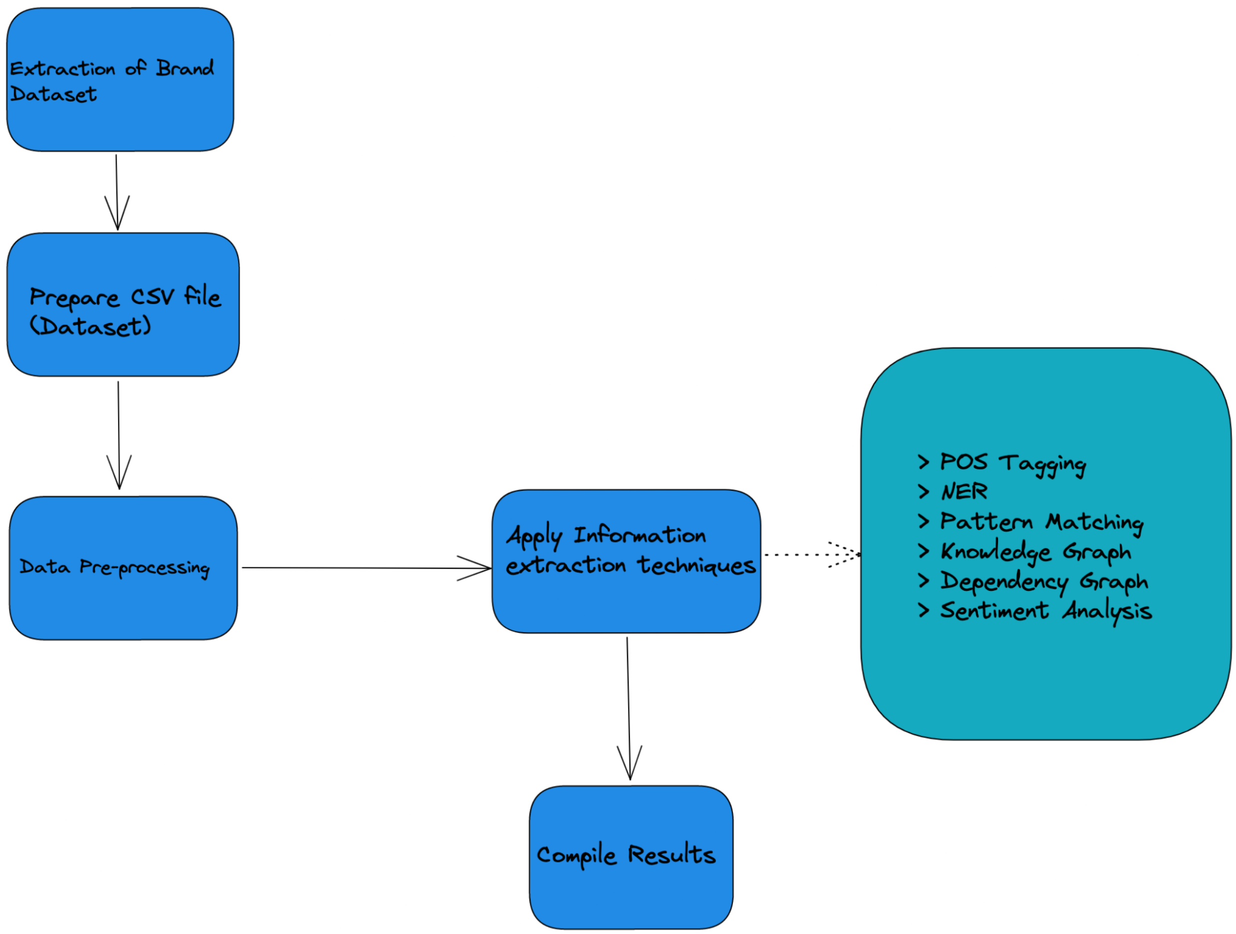
Architecture diagram





Flow diagram (According to current progress)

We plan to extract data directly from twitter as a further improvement to present real-time analysis



Pseudo code

Extracting tweets from twitter through tweepy library and storing the results into a csv file .
Here we search for tweets with specific keywords associated with a particular car brand/ Model.

```
def extract_tweets(search_words):

   consumer_key = os.getenv['consumer_key']
   consumer_secret = os.getenv['consumer_secret']
   access_key= os.getenv['access_key']
   access_secret = os.getenv['access_secret']

   auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
```

```
auth.set_access_token(access_key, access_secret)

api = tweepy.API(auth,wait_on_rate_limit=True)

csvFile = open('./data/tweets.csv', 'a')
csvWriter = csv.writer(csvFile)

search_words = search_words
new_search = search_words + " -filter:retweets"

for tweet in tweepy.Cursor(api.search_tweets,q=new_search,count=10,
                        lang="en",
                        since_id=0).items():
    csvWriter.writerow([tweet.created_at,
    tweet.text.encode('utf-8'),
    tweet.user.screen_name.encode('utf-8'),
    tweet.user.location.encode('utf-8')])
```

Data Pre-processing

```
def clean(text):

   # removing paragraph numbers
   text = re.sub('[0-9]+.\t','',str(text))
   # removing new line characters
   text = re.sub('\n ','',str(text))
   text = re.sub('\n',' ',str(text))
   # removing apostrophes
   text = re.sub("'s",'',str(text))
   # removing hyphens
   text = re.sub("-",' ',str(text))
   text = re.sub("— ",'',str(text))
   # removing quotation marks
   text = re.sub('\"','',str(text))
   # removing salutations
   text = re.sub("Mr\.",'Mr',str(text))
   text = re.sub("Mrs\.",'Mrs',str(text))
   # removing any reference to outside text
```

```
   text = re.sub("[\(\[].*?[\)\]]", "", str(text))

   return text
```

Apart from these techniques we also apply tokenization, stemming and lemmatiztion (performed by Spacy library under the hood)

We wrap all our Information Extraction approaches into a single function which takes in the parameters of the name of the automobile brand

```
from spacy.symbols import nsubj, VERB, ADJ

def IE_Operations(review):
 # create spacy doc
 doc = nlp(review)
 adjectives = set()
 verbs_all = set()
 # applying POS to each token
 for token in doc:
     if token.pos_ not in ["SPACE", "DET", "ADP", "PUNCT", "AUX", "SCONJ",
"CCONJ", "PART"]:
       print(token.text,'->',token.pos_)
     if(token.pos_=="ADJ"):
       adjectives.add(token.text)
     if(token.pos_=="VERB"):
       verbs_all.add(token.text)
 displacy.render(doc, style='dep',jupyter=True)
 # Finding a verb with a subject
 verbs = set()
 for possible_subject in doc:
       if possible_subject.dep == nsubj and possible_subject.head.pos ==
VERB:
         verbs.add(possible_subject.head)
 print(verbs)
 # Finding adjectives with a subject
 print(adjectives)
 # NER
 for ent in doc.ents:
   print(ent.text, ent.start_char, ent.end_char, ent.label_)
 knowledge_graph(review)

def IE_brand(brand):
```

```
path = "/content/Scraped_Car_Review_" + brand + ".csv"
df = pd.read_csv(path,delimiter=',', nrows = 100)
df['Review_clean'] = df['Review'].apply(clean)
df['Review_clean'][2]
reviews = df['Review_clean'][0:10]
reviews = np.array(reviews)
for review in reviews:
  IE_Operations(review)
```

Experiment and Results

Dataset

We worked with the edmundsconsumer-car-ratings-and-reviews dataset which has 62 different sub-datasets for popular car brands with customer reviews for the vehicles they bought from that specific company.

This dataset contains data about customer thoughts, car models and ratings given after each purchase.

Currently, this dataset has data of 62 major brands such as Acura, AlfaRomeo ,AMGeneral, Aston Martin, Audi, Bentley, and others

Working with the Dataset

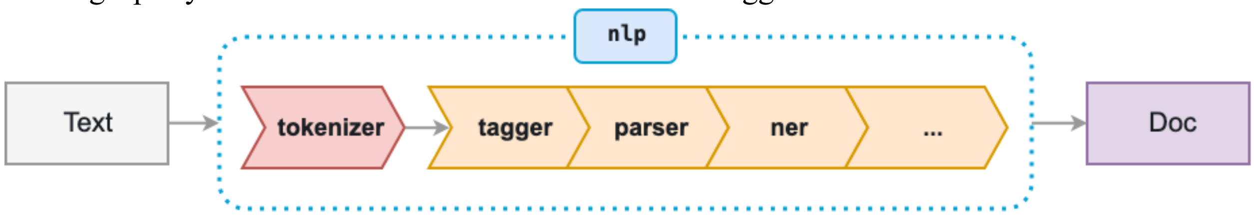
The User inputs the car brand name based on which the relevant sub-dataset is chosen and worked on. Once the correct sub-dataset is chosen, we create a dataframe from the csv file using Pandas library and perform text pre-procesing to remove unnecessary words, punctuations etc, which contribute little or nothing to the overall analysis.

Using the Kaggle API we add the dataset into our Collab notebook and unzip all 62 individual csv files. From the relevant dataset we extract only the top 100 rows for the purpose of experimentation.
Next we create a list of all pre-processesd sentences from the dataframe after selecting the 'Reviews' column which stores customer reviews for a vehicle bought from the chosen automobile brand.

We primarily utilize the SpaCy library in Python for the purpose of our experiment simulation

It's tough to process raw text effectively since most words are uncommon, and it's typical for words that appear to be entirely different to imply almost the same thing. The same words in a different sequence might have entirely distinct meanings. In many languages, even dividing text into useable word-like units can be challenging.

While certain issues may be solved using simply raw letters, it is typically preferable to employ linguistic understanding to provide important information. That's precisely what spaCy is supposed to do: take in raw text and return a Doc object with various annotations.
spaCy is one of the best text analysis library. spaCy excels at large-scale information extraction tasks and is one of the fastest in the world. It is also the best way to prepare text for deep learning. spaCy is much faster and accurate than NLTKTagger and TextBlob.



SpaCy's Processing Pipeline

Methodology

Once we have acquired the list of reviews, we perform the following mentioned operations on each review:

- ❖ POS tagging

  Here we utilize the SpaCy library to load a pre-trained NLP model called 'en_core_web_sm' using which predicts the correct Part of Speech of each word in the review sentence

  SpaCy can parse and tag a Doc after it has been tokenized. This is where spaCy's trained pipeline and statistical models come in, allowing it to anticipate which tag or label is most likely to apply in this situation. A trained component contains binary data that is generated by giving a system enough instances for it to make language-specific predictions — for example, in English, a word after "the" is most often a noun.

- ❖ Creating Dependency Graph

  Visualizing a dependency parse or named entities in a text isn't only a fun NLP demonstration, it can also assist you to speed up the creation and debugging of your code and training.

  Using the Dependency graph we unearth relations and which can be useful for further analysis

- ❖ Creating Knowledge Graph

We produce a knowledge graph after using NLP techniques such as sentence segmentation, dependency parsing, parts of speech tagging, and entity recognition.
The first step in building a knowledge graph is to split the text document or article into sentences. Then, we will shortlist only those sentences in which there is exactly 1 subject and 1 object

It is not difficult to extract a single word unit from a phrase. With the use of parts of speech (POS) tags, we can easily do this. Our entities would be the nouns and proper nouns.

We need edges to link the nodes (entities) in order to form a knowledge graph. These edges represent the connections between two nodes.

- Filtering out verbs and adjectives to perform sentiment analysis

We do this by simply creating an 'if' condition in Python, which checks if the POS tag is a VERB or an ADJ and then adds it to a set.

- Named Entity Recognition

NER locates and categorizes identified entities included in unstructured text into conventional categories such as person names, locations, organizations, time expressions, amounts, monetary values, percentages, codes, and so on. Spacy includes a statistical entity recognition system that gives labels to tokens in continuous spans.

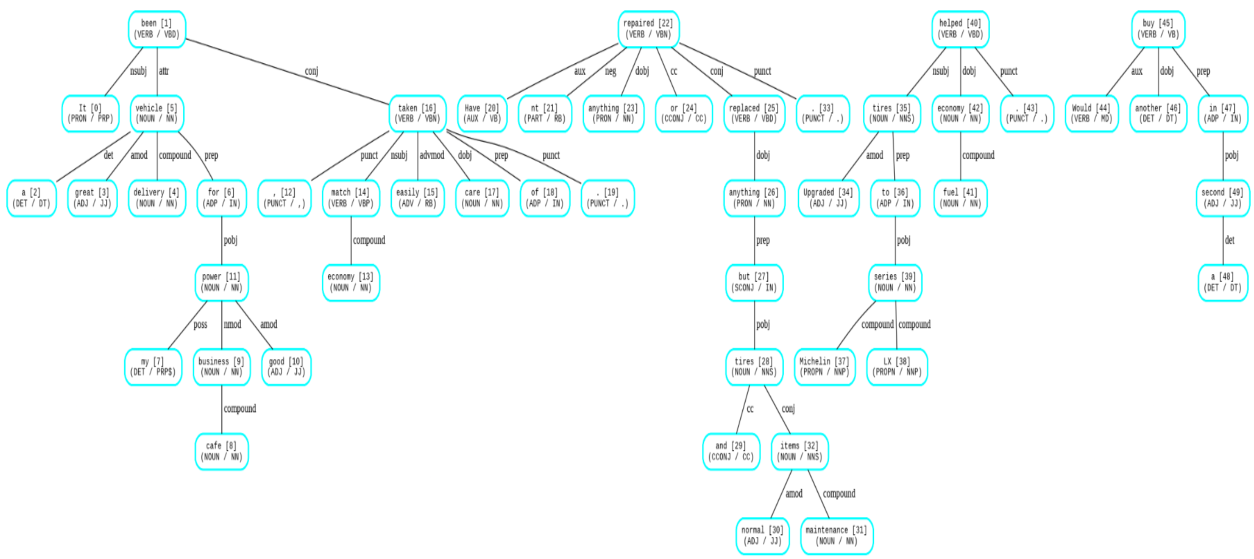
The 'ner' pipeline component in Spacy finds token spans that match a predefined collection of named entities. These may be found in a Doc object's 'ents' attribute.

Results

For the given sentence (taken from the dataset after pre-processing) we perform POS Tagging and creating dependency Graph

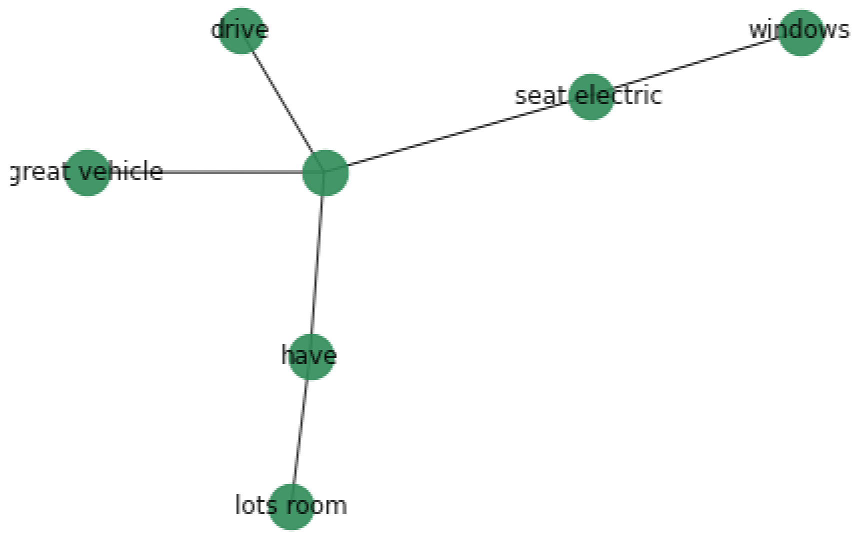
"It been a great delivery vehicle for my cafe business good power, economy match easily taken care of. Haven't repaired anything or replaced anything but tires and normal maintenance items. Upgraded tires to Michelin LX series helped fuel economy. Would buy another in a second"

```
It -> PRON
been -> VERB
great -> ADJ
delivery -> NOUN
vehicle -> NOUN
cafe -> NOUN
business -> NOUN
good -> ADJ
power -> NOUN
economy -> NOUN
match -> VERB
easily -> ADV
taken -> VERB
care -> NOUN
repaired -> VERB
anything -> PRON
replaced -> VERB
anything -> PRON
tires -> NOUN
normal -> ADJ
maintenance -> NOUN
items -> NOUN
Upgraded -> ADJ
tires -> NOUN
Michelin -> PROPN
LX -> PROPN
series -> NOUN
helped -> VERB
fuel -> NOUN
economy -> NOUN
Would -> VERB
buy -> VERB
second -> ADJ
***************************
```



Knowledge graph for the following sentence looks as follows

"Great work vehicle. Drives nice. has lots of room. Easy to handle, bucket seats electric windows"



“ So, I just bought my first used car. It's a 1999 Jeep Cherokee. I've read Kelley's Blue Book, and the price was right on target. It has 179,000 miles on it. Jeep owners tell me it will still go on and on. I've also read most of these reviews, and they all seem very positive..The man who sold it told me it just needs a muffler; it is a bit loud. Other than that, it rode well. One owner. It's my first Jeep, so I'm keeping my fingers crossed. I'll be very happy if I can get at least a year out of it and, hopefully, not spend too much money on it for repairs. But time will tell.”

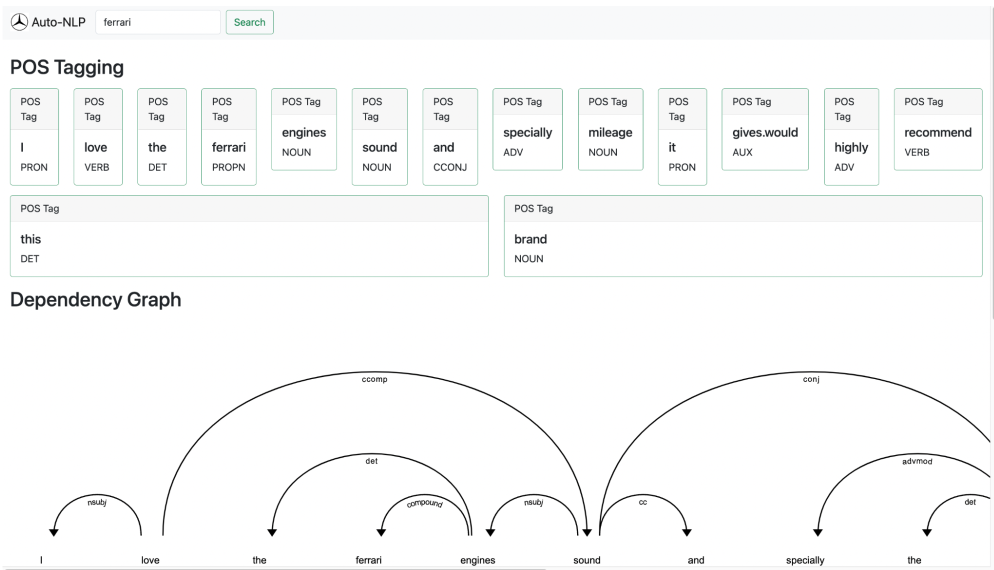
```
{told, tell, read, read, rode, keeping, seem, bought, sold, needs, go, tell}
************************************************************

{'much', 'right', 'loud', 'Other', 'first', 'happy', 'most', 'positive'}
************************************************************

first 22 27 ORDINAL
1999 44 48 DATE
Jeep Cherokee 49 62 ORG
Kelley Blue Book 75 91 ORG
179,000 miles 136 149 QUANTITY
Jeep 158 162 ORG
One 383 386 CARDINAL
first 401 406 ORDINAL
Jeep 407 411 ORG
at least a year 481 496 DATE
************************************************************
```

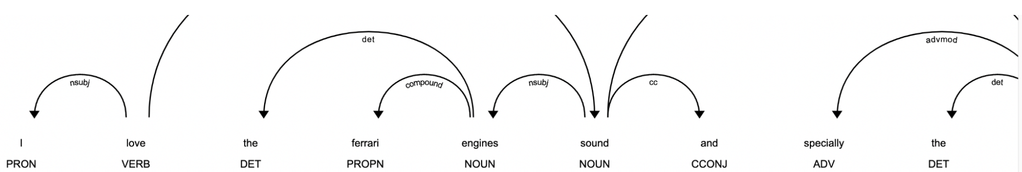
Web Application Screenshots

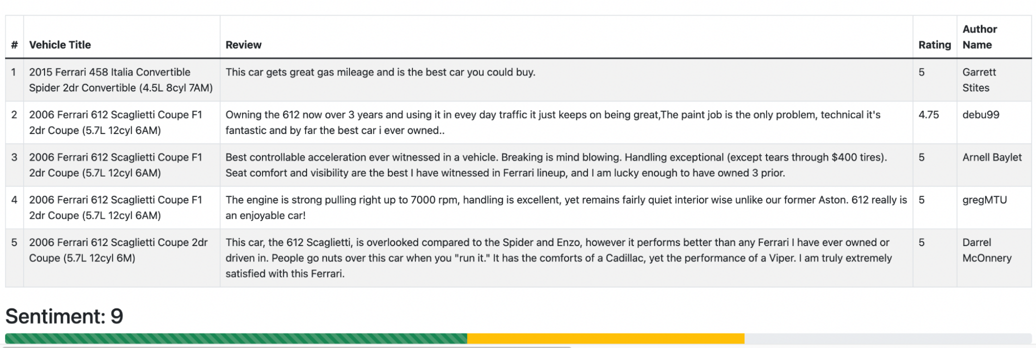
Github Link: https://github.com/batman004/Auto-Utility

We transformed our NLP-based algorithms to RESTfull APIs and expose them as endpoints that can be consumed by our frontend.

We have developed a comprehensive dashboard that sends out requests to our backend API, which responds with real-time data associated with a car brand.







#	Vehicle Title	Review	Rating	Author Name
1	2015 Ferrari 458 Italia Convertible Spider 2dr Convertible (4.5L 8cyl 7AM)	This car gets great gas mileage and is the best car you could buy.	5	Garrett Stites
2	2006 Ferrari 612 Scaglietti Coupe F1 2dr Coupe (5.7L 12cyl 6AM)	Owning the 612 now over 3 years and using it in evey day traffic it just keeps on being great,The paint job is the only problem, technical it's fantastic and by far the best car i ever owned..	4.75	debu99
3	2006 Ferrari 612 Scaglietti Coupe F1 2dr Coupe (5.7L 12cyl 6AM)	Best controllable acceleration ever witnessed in a vehicle. Breaking is mind blowing. Handling exceptional (except tears through $400 tires). Seat comfort and visibility are the best I have witnessed in Ferrari lineup, and I am lucky enough to have owned 3 prior.	5	Arnell Baylet
4	2006 Ferrari 612 Scaglietti Coupe F1 2dr Coupe (5.7L 12cyl 6AM)	The engine is strong pulling right up to 7000 rpm, handling is excellent, yet remains fairly quiet interior wise unlike our former Aston. 612 really is an enjoyable car!	5	gregMTU
5	2006 Ferrari 612 Scaglietti Coupe 2dr Coupe (5.7L 12cyl 6M)	This car, the 612 Scaglietti, is overlooked compared to the Spider and Enzo, however it performs better than any Ferrari I have ever owned or driven in. People go nuts over this car when you "run it." It has the comforts of a Cadillac, yet the performance of a Viper. I am truly extremely satisfied with this Ferrari.	5	Darrel McOnnery

Conclusion

Information extraction is not a simple NLP operation to do. To better comprehend the data's structure and what it has to give, you must spend time with it. we applied theoretical knowledge to real-world situations. We attempted to extract information from a text collection using typical information extraction techniques. To extract information from the text, we looked for significant terms and connections in the text data. To extract useful data, this technique needs a combination of computing power and human labor.

Using the given dataset, we successfully applied well-known information extraction techniques such as POS tagging, NER, Dependency graph, etc to extract valuable insights from the customer reviews given by people who bought products from a given brand.
This information can be of immense value to automobiles companies aiming to launch a new product into the market or trying to analyze the existing performance of a product already in the market to judge its success in terms of customer satisfaction.

We further aim to add other information extraction techniques such a topic classification, Sentiment Analysis, Text Summarization, Aspect Mining and Topic Modeling to strengthen our results and provide better actionable insights to brands. Also we plan to present all the information extracted in a well layed-out dashboard, where even a person without much knowledge of NLP can gauge the data and make inferences.

References

[1] Minaee, Shervin, et al. "Deep learning-based text classification: a comprehensive review." *ACM Computing Surveys (CSUR)* 54.3 (2021): 1-40.
[2] Li, Jing, et al. "A survey on deep learning for named entity recognition." *IEEE Transactions on Knowledge and Data Engineering* 34.1 (2020): 50-70.
[3] Opdahl, Andreas L. "Knowledge Graphs and Natural-Language Processing." *Big Data in Emergency Management: Exploitation Techniques for Social and Mobile Data*. Springer, Cham, 2020. 75-91.
[4] Perera, Nadeesha, Matthias Dehmer, and Frank Emmert-Streib. "Named entity recognition and relation detection for biomedical information extraction." *Frontiers in cell and developmental biology* (2020): 673.
[5] Luo, Junwei, and Xiao Xue. "Research on information retrieval system based on Semantic Web and multi-agent." *2010 International Conference on Intelligent Computing and Cognitive Informatics*. IEEE, 2010.
[6] Singh, Sonit. "Natural language processing for information extraction." *arXiv preprint arXiv:1807.02383* (2018).
[7] Young, Tom, et al. "Recent trends in deep learning-based natural language processing." *ieee Computational intelligenCe magazine* 13.3 (2018): 55-75.
[8] Gloutnikov, Stefan S. "Using Sentiment Analysis and Pattern Matching to Signal User Review Abnormalities." (2018).
[9] Pei, Wenzhe, Tao Ge, and Baobao Chang. "An effective neural network model for graph-based dependency parsing." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2015.
[10] Sarkar, Kamal. "Part-of-speech tagging for code-mixed indian social media text at icon 2015." *arXiv preprint arXiv:1601.01195* (2016).
[11] Nastase, Vivi, Rada Mihalcea, and Dragomir R. Radev. "A survey of graphs in natural language processing." *Natural Language Engineering* 21.5 (2015): 665-698.
[12] Bhatia, Parul Kalra, Tanya Mathur, and Tanaya Gupta. "Survey paper on information retrieval algorithms and personalized information retrieval concept." *International Journal of Computer Applications* 66.6 (2013).
[13] Russell-Rose, Tony, and Mark Stevenson. "The role of natural language processing in information retrieval: Searching for meaning and structure." (2009): 215-231.
[14] Lopatenko, Andrei. "Information retrieval in current research information systems." *arXiv preprint cs/0110026* (2001).
[15] Lewis, David D., and Karen Spärck Jones. "Natural language processing for information retrieval." *Communications of the ACM* 39.1 (1996): 92-101.
[16] Barbero, Cristina, and Vincenzo Lombardo. "Dependency graphs in natural language processing." *Congress of the Italian Association for Artificial Intelligence*. Springer, Berlin, Heidelberg, 1995.