Employee Attrition

Abstract: Human Resource is vital role for bringing most talented within company, employee attrition really pays huge cost to company. As we see ML and NLP is Disrupting Technology so we are using ML Boosting algorithms for Attrition prediction from already present datasets and NLP state of arts methods to analyze various dimensions of segment of HR workflows such as feedback, Resume Screening to extract behavior, intents and emotions from employees and classify or rank the resumes. We covered literature surveys from pre used methodologies to makes important decision for feature selection, NLP modules to makes use of and Knowledge graphs to visualize clear picture for making decision stronger.

1. Introduction:

As boosting Algorithm usually outperforms well on Multiple datatypes, as well as catboost is capable of handling feature engineering process in hand. To extract information from usually requires OCR capabilities which can be deduced with tesseract or any predefined library, again it comes with various technical challenges which can be overcome

NLP gives essence to people function and turns out to be most powerful techniques to analyze various capabilities of Human behavior, emotions, sentiments, analyze various heterogenous relationship with structuring and visualization makes vital steps. Knowledge Graph to get various relationships from entities with node and score as strength for relationship.

2. Literature Survey

2.1 Attrition [1]

Employee attrition is the reduction of staff when employees leave the organization without emphasizing on their replacement.

Attrition Rate = (No. of employees resigned/No. of employees at the start of the month + no. of employees joined - no. of employees resigned) **x 100**

2.1.1 What Is Influencing Your Attrition Rate?

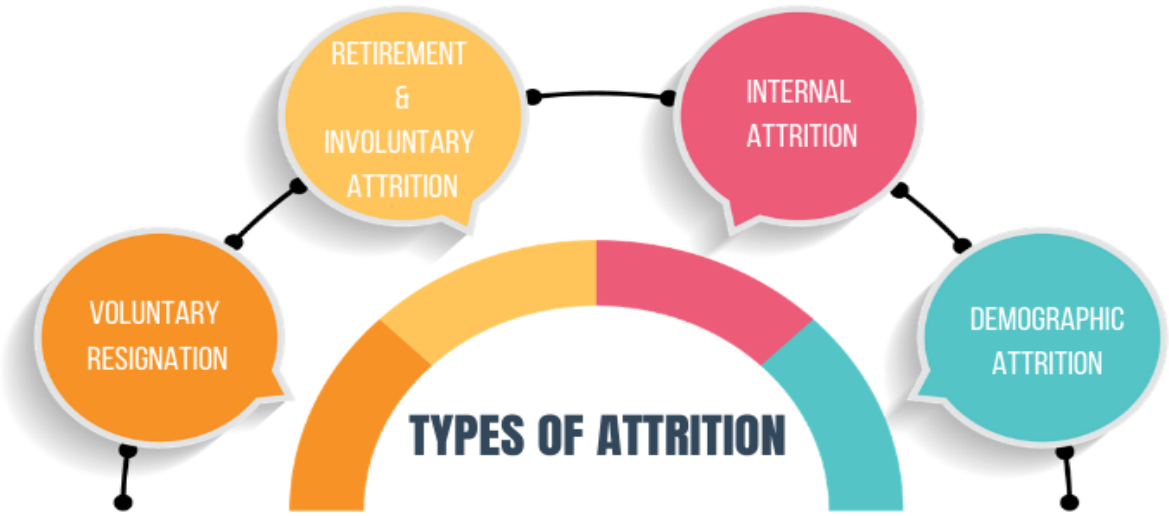
When they leave it's helpful to ask questions about?

- Managers
- Team atmosphere
- Job characteristics
- Pay and rewards
- Stress and workload
- Workforce demographics

2.1.1 Types of Attrition

Attrition can be of different types and they include-

- Voluntary Resignation
- Involuntary Attrition
- Retirement
- Internal Attrition
- Demographic.



2.1.2 Getting Beyond the Common Misconceptions about Attrition[2]

The most common misconceptions about attrition are the following:

- Attrition is only related to what you do or don't do well as a manager or company
- People leave for only a single reason
- All attrition is the same
- All attrition should be prevented
- You can compare the attrition count of one group directly to another
- You can control employee attrition with global, one-size-fits-all efforts

2.1.3 Questions to Tackle[3]

1. Can I predict which staff members will leave next?
2. Can I create a system which may assist in staff retention?

3.Data for Attrition

Data is oil for ML problems so analysis of data is prerequisites task, to get most features with relationships to get more depth. Dataset is collected from various resources[7][4][3][6].

3.1 IBM HR Analytics Employee Attrition & Performance

This data set presents an employee survey from IBM, indicating if there is attrition or not. The data set contains approximately 1500 entries. Given the limited size of the data set, the model should only be expected to provide modest improvement in identification of attrition vs a random allocation of probability of attrition [4].

3.1 CHECK TABLE

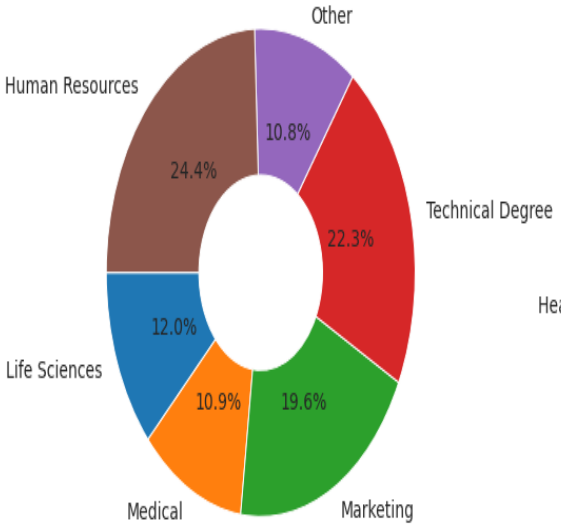
Name	Description
AGE	Numerical Value
ATTRITION	Employee leaving the company (0=no, 1=yes)
BUSINESS TRAVEL	(1=No Travel, 2=Travel Frequently, 3=Travel Rarely)
DAILY RATE	Numerical Value - Salary Level
DEPARTMENT	(1=HR, 2=R&D, 3=Sales)
DISTANCE FROM HOME	Numerical Value - THE DISTANCE FROM WORK TO HOME
EDUCATION	Numerical Value
EDUCATION FIELD	(1=HR, 2=LIFE SCIENCES, 3=MARKETING, 4=MEDICAL SCIENCES, 5=OTHERS, 6= TE

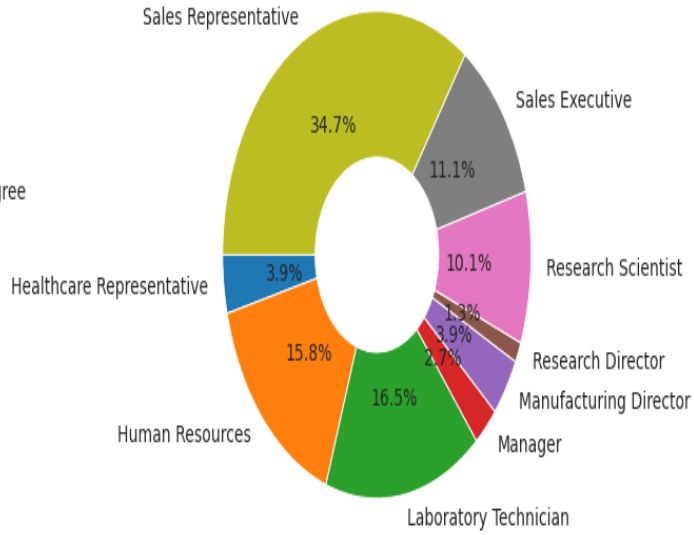
Name	Description
EMPLOYEE COUNT	Numerical Value
EMPLOYEE NUMBER	Numerical Value - EMPLOYEE ID
ENVIROMENT SATISFACTION	Numerical Value - SATISFACTION WITH THE ENVIROMENT
GENDER	(1=FEMALE, 2=MALE)
HOURLY RATE	Numerical Value - HOURLY SALARY
JOB INVOLVEMENT	Numerical Value - JOB INVOLVEMENT
JOB LEVEL	Numerical Value - LEVEL OF JOB
JOB ROLE	(1=HC REP, 2=HR, 3=LAB TECHNICIAN, 4=MANAGER, 5= MANAGING DIRECTOR, 6= REASEARCH DIRECTOR, 7= RESEARCH SCIENTIST, 8=SALES EXECUTIEVE, 9= SALE REPRESENTATIVE)
JOB SATISFACTION	Numerical Value - SATISFACTION WITH THE JOB
MARITAL STATUS	(1=DIVORCED, 2=MARRIED, 3=SINGLE)
MONTHLY INCOME	Numerical Value - MONTHLY SALARY
MONTHY RATE	Numerical Value - MONTHY RATE
NUMCOMPANIES WORKED	Numerical Value - NO. OF COMPANIES WORKED AT
OVER 18	(1=YES, 2=NO)

Name	Description
OVERTIME	(1=NO, 2=YES)
PERCENT SALARY HIKE	Numerical Value - PERCENTAGE INCREASE IN SALARY
PERFORMANCE RATING	Numerical Value - ERFORMANCE RATING
RELATIONS SATISFACTION	Numerical Value - RELATIONS SATISFACTION
STANDARD HOURS	Numerical Value - STANDARD HOURS
STOCK OPTIONS LEVEL	Numerical Value - STOCK OPTIONS
TOTAL WORKING YEARS	Numerical Value - TOTAL YEARS WORKED
TRAINING TIMES LAST YEAR	Numerical Value - HOURS SPENT TRAINING
WORK LIFE BALANCE	Numerical Value - TIME SPENT BEWTWEEN WORK AND OUTSIDE
YEARS AT COMPANY	Numerical Value - TOTAL NUMBER OF YEARS AT THE COMPNAY
YEARS IN CURRENT ROLE	Numerical Value -YEARS IN CURRENT ROLE
YEARS SINCE LAST PROMOTION	Numerical Value - LAST PROMOTION
YEARS WITH CURRENT MANAGER	Numerical Value - YEARS SPENT WITH CURRENT MANAGER

3.1.1 Analysis and Summary Glance

Q1: Which department and job role has seen the most attrition?[5]





3.2 **Solving Staff Attrition with Data[3]**

The dataset is simulated and contained the following fields:

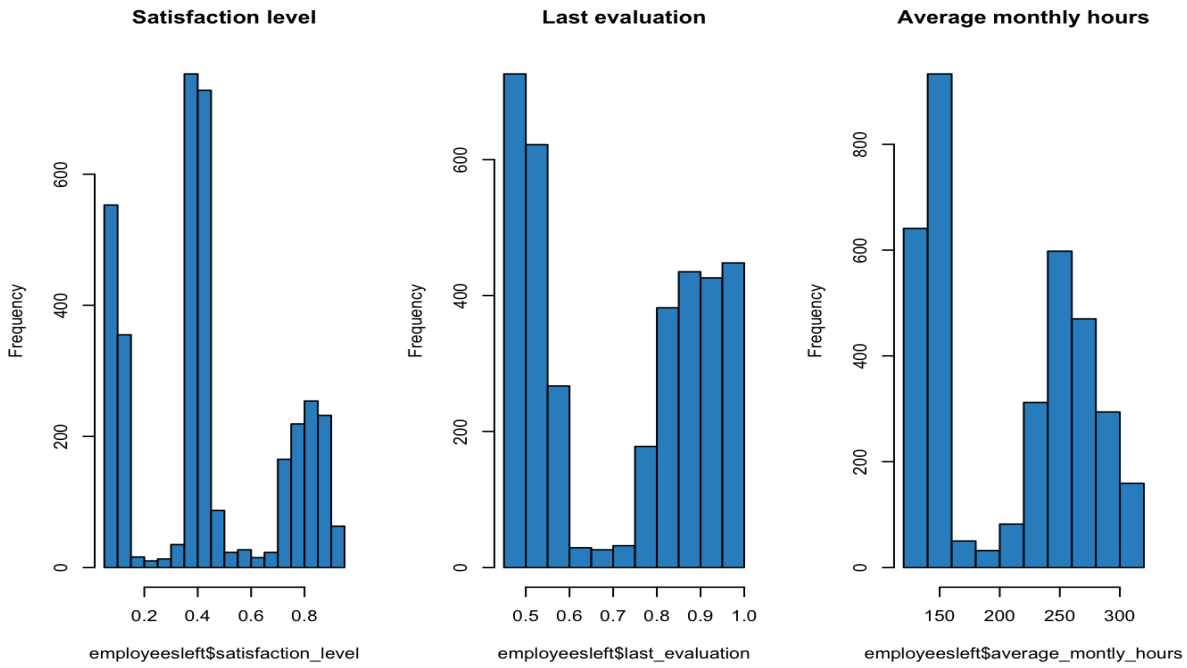
- Employee satisfaction level
- Last evaluation
- Number of projects
- Average monthly hours
- Time spent at the company
- Whether they have had a work accident
- Whether they have had a promotion in the last 5 years

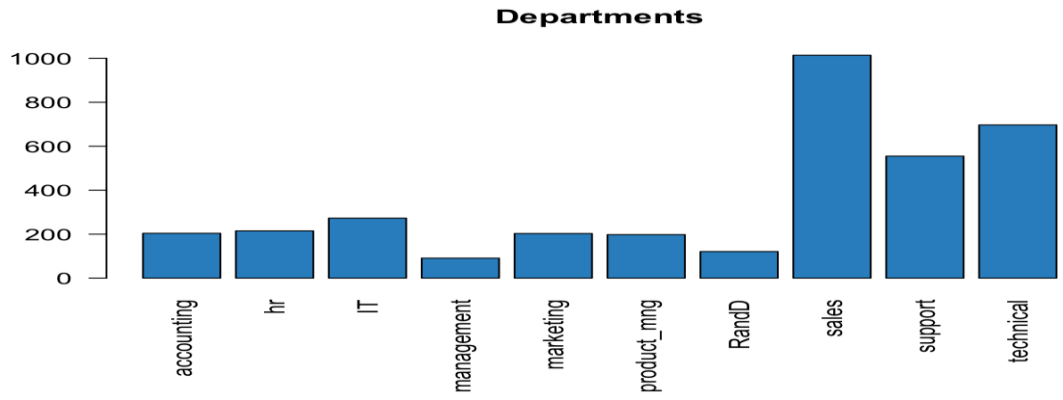
- Department
- Salary
- Whether the employee has left

3.2.1 Formulation of problem[3].

- **Descriptive Analytics:** What are the observations that help us shape our various hypotheses about staff attrition?
- **Predictive Analytics:** Which members of staff are about to leave?
- **Prescriptive Analytics:** What insights or suggestions can be made concerning those members of staff who are likely to leave?

3.2.2 Graph and Summary [6]





4.0 NLP for Attrition

Natural language processing is an ever-growing interest area in the analytics application spectrum and is relevant to HR. In fact, it can revolutionize the quality of insights. In this article, we will explain you how[8].

4.1 The Role of Natural Language Processing in HR[9]

1. **NLP in Recruitment**
2. **NLP in Employee Engagement**
3. **NLP in Employee Social Media Analytics**

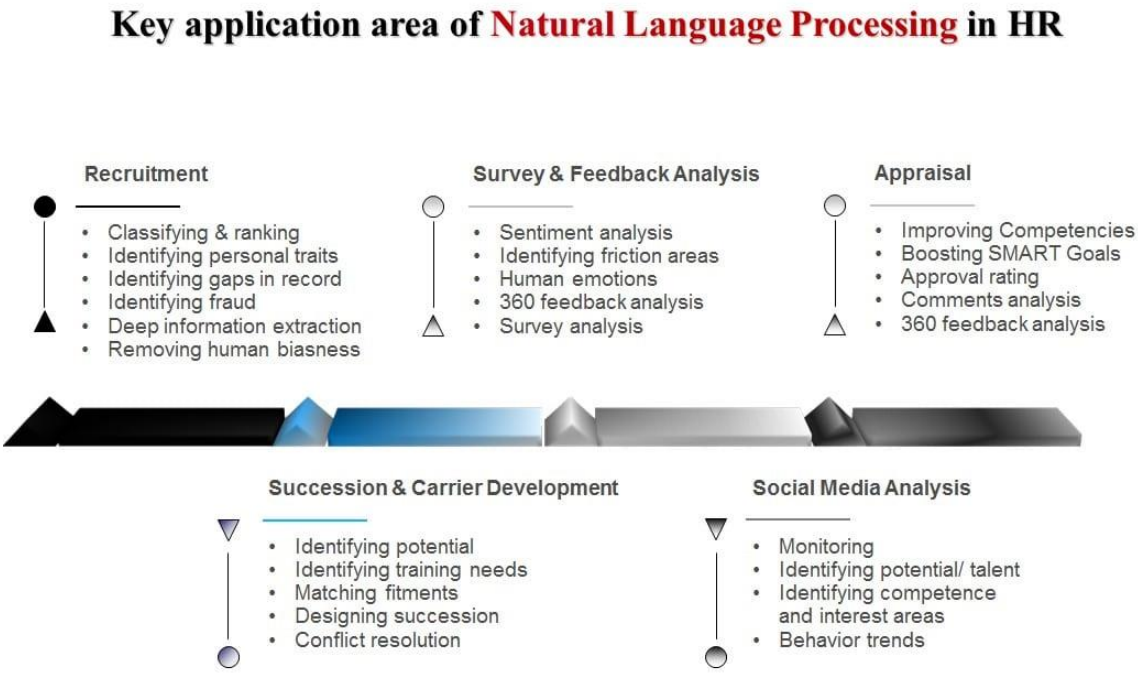
4.3 Natural Language Processing Examples in HR[10]

- **NLP in Recruiting**
- **NLP in Interviewing**
- **NLP in Onboarding and Training**
- **NLP in Employee Engagement**
- **NLP in Behavioral Training**

4.4 NLP Application in HR [11].

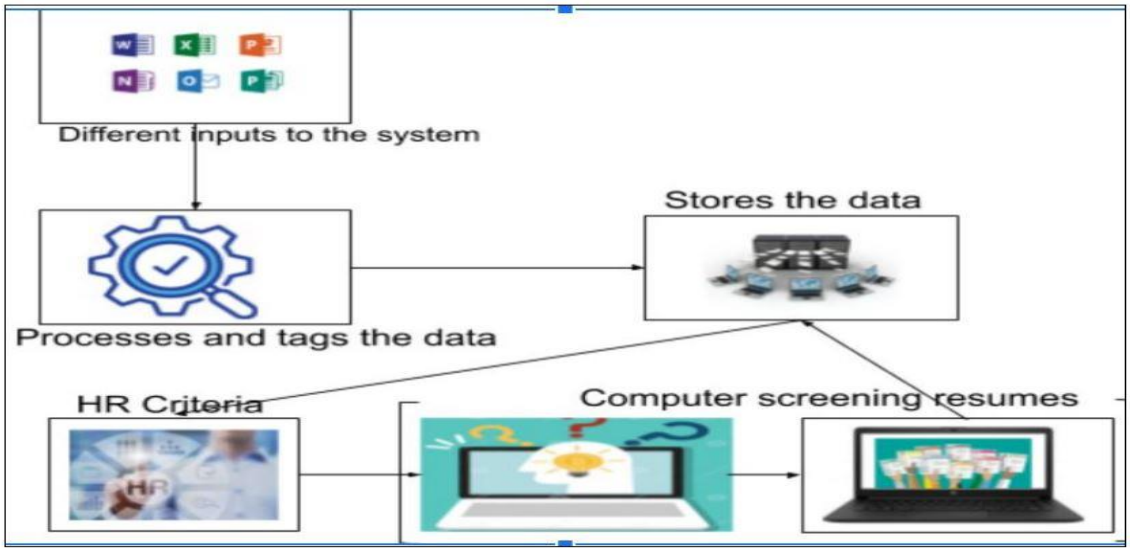
- **1- NLP & Voice of Employee**
- **2- The NLP for recruitment assistance**
- **3- HR Chatbot**

4.4 Key Application for HR[8]

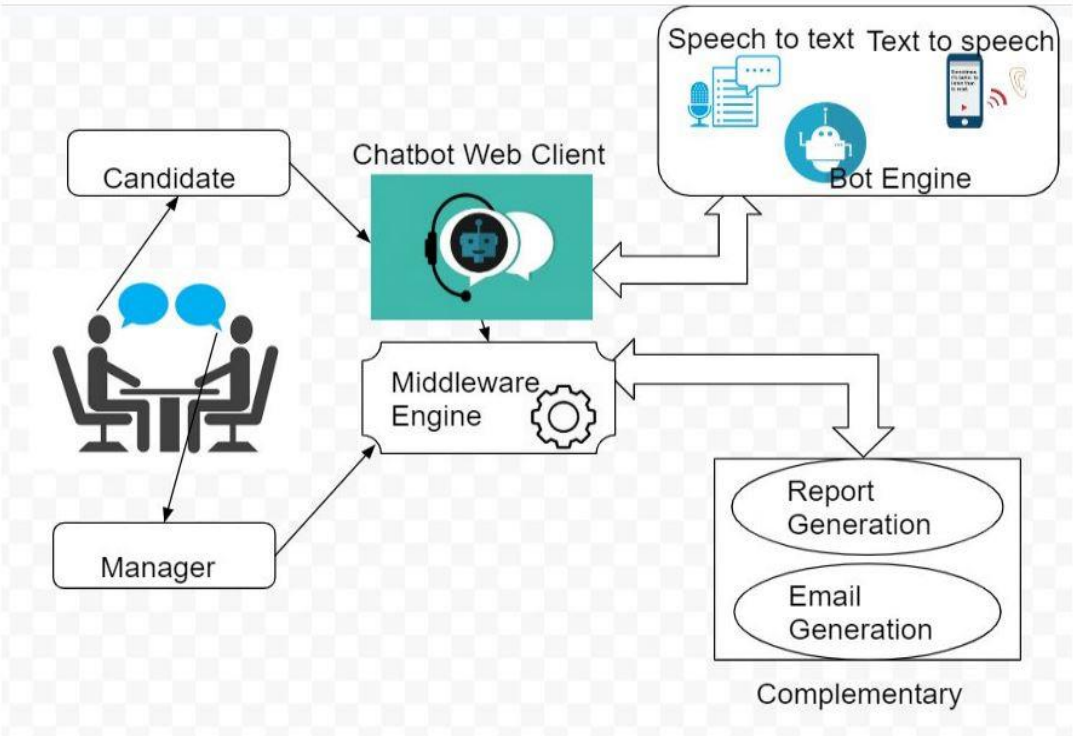


5.1 Built Methodology

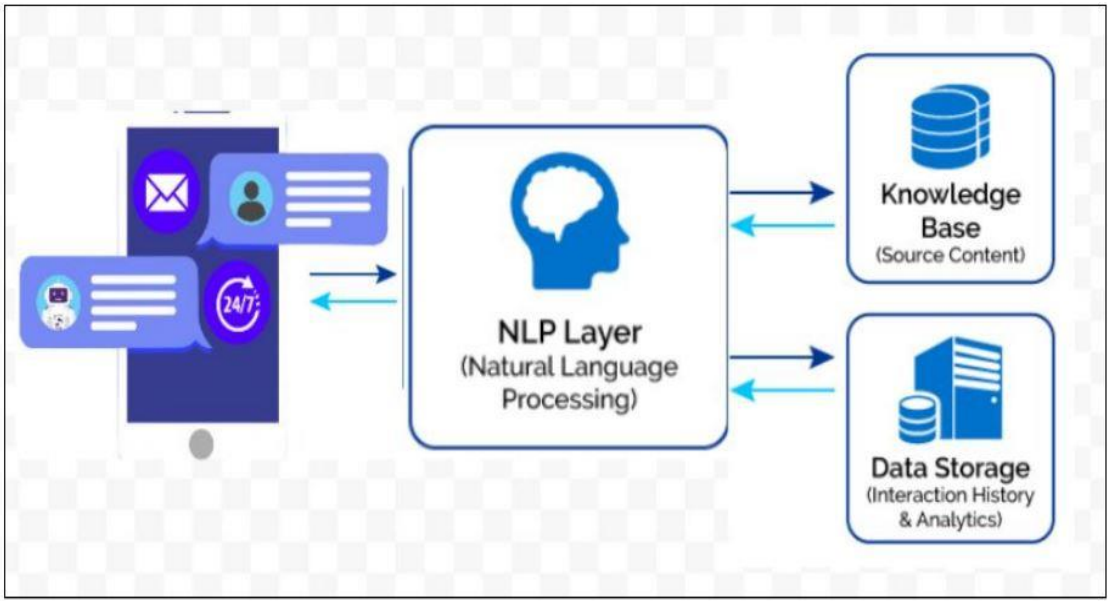
- **NLP for Resume Screening:**



- **NLP for Interviewing**



- **Chatbots for Employees**



Methodology:

NLP Module:

Feedback is often playing important role for analyzing various characteristics of employee such as emotions, intentions, mentality, behavior and sentiments which can be detected from language modeling. Which can be discovered from NLP. NLP modules used starts from preprocessing unit were removing noise and lowercasing with lemmatizing often challenging and time consuming. Using various tokenization makes decision making parameter for feeding text classification module so going white space to sentence and word piece tokenization covers lot our task to define

Resume Ranker:

The recruiter will be helped by the procedure mentioned to shortlist the resumes based on the core skills and previous experiences. Here are the steps involving while resume screening. The below illustration shows major steps involved in the process of resume screening. It does the data gathering in the initial step - resumes are gathered from different sources and data will be pushed to Data processing step. Data will be converted to one common format when pushed to Data processing and tagging. The second step involves Data Processing - it extracts the required information from the provided source. Following this, there is an Attribute Tagger. In this step, the main attributes like candidate name, experience, skills will be checked. Here, the significant tool of NLP - Named Entity Recognition (NER) is also used as it identifies the proper nouns from an open domain text. The required and necessary information is stored in the storage for future usage.

Feature Feeder:

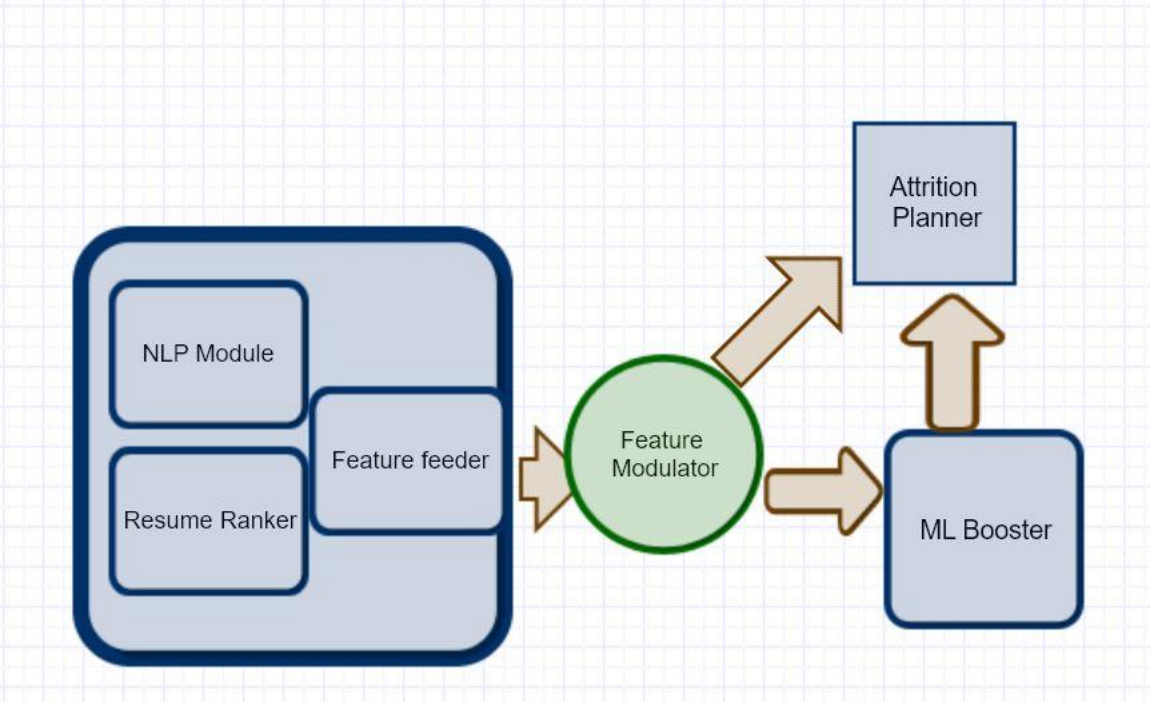
We have various features from various that play important role for attrition to retention which makes decision consideration. Feature engineering and correlation components to relates our features. After smashing unwanted features, it output is feeded in catboost.

ML Booster:

Boosting algorithm is often beats most ML algorithm so using catboost for categorizing multiple datatypes, missing values, handling outlier often taken care with parameters and Hyperparameter tuning makes stronger decisions.

Attrition Planer

Final stage where knowledge graph is used to visualize employee department attrition.



Design Module in Nutshell

References:

1. **https://blog.vantagecircle.com/employee-attrition/**
2. **https://www.linkedin.com/pulse/analyzing-employee-attrition-mike-west/**
3. **https://towardsdatascience.com/solving-staff-attrition-with-data-3f09af2694cd**
4. http://inseaddataanalytics.github.io/INSEADAnalytics/groupprojects/January2018FBL/IBM_Attrition_VSS.html
5. https://medium.com/jovianml/ibm-hr-analytics-data-eda-35728dec9d84
6. http://inseaddataanalytics.github.io/INSEADAnalytics/groupprojects/January2017/AnalysisofHR%5B1%5D.html
7. https://www.analyticsinhr.com/blog/hr-data-sets-people-analytics/
8. https://www.analyticsinhr.com/blog/natural-language-processing-revolutionize-human-resources/
9. https://www.hrtechnologist.com/articles/digital-transformation/3-ways-natural-language-processing-nlp-will-transform-hr-in-2019/#:~:text=NLP%20provides%20a%20data%2Ddriven,traits%2C%20and%20eliminate%20human%20biases.
10. https://www.plugandplaytechcenter.com/resources/introduction-nlp-and-how-its-changing-future-hr/
11. https://www.proxem.com/en/nlp-applications-for-human-resources/#2
12. https://www.profinda.com/en/machine-learning-in-talent-acquistion/