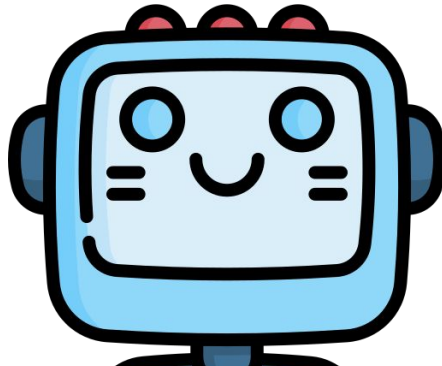


# Find Technical Words



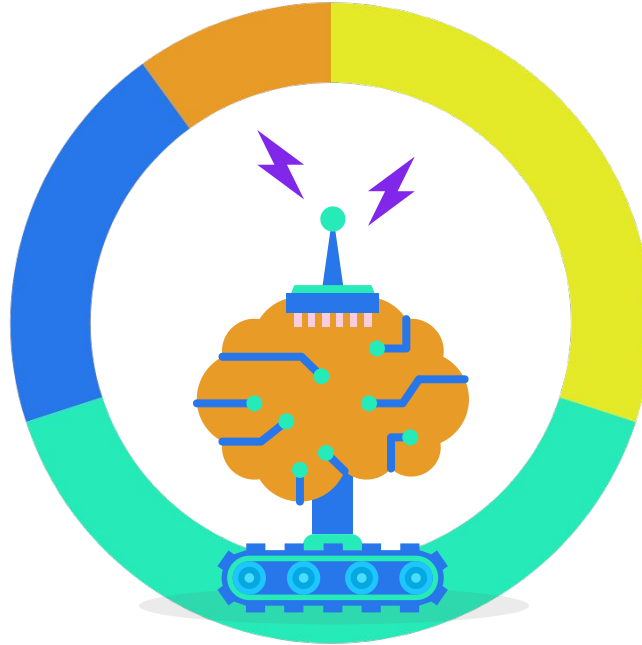
# Data Analytics

**30%**

**Repeat Words**

**5-7%**

**STOPWORDS**



**2%**

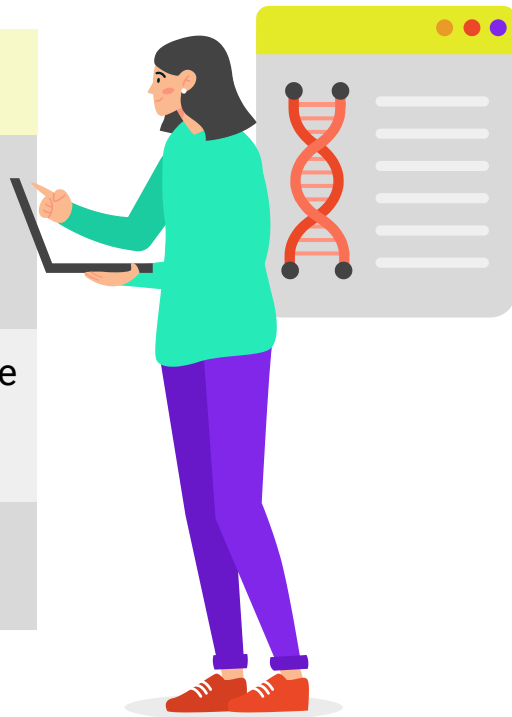
**Missing Values**

**15677**

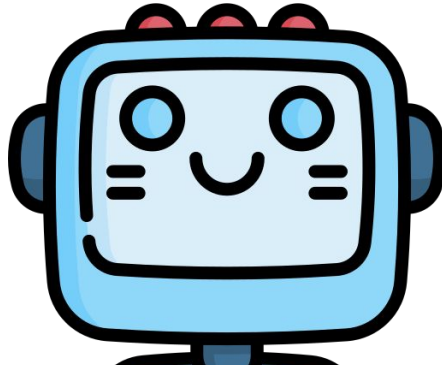
**UNIQUE WORDS**

# Models We Tried

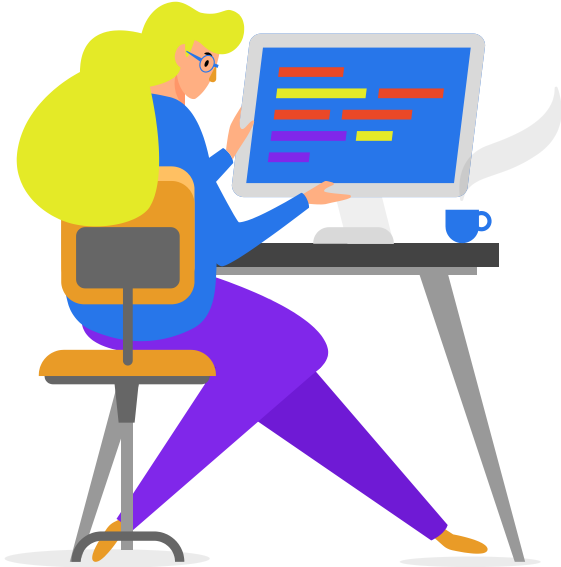
Models/Algorithms	
SPACY	A pre-trained model used to find similarities and then remove non technical words!
Cosine Similarity	Calculates the vector angle between the 2 words and thus we get a rank between 1 & 0
Word2vec	Find similar words to get the technical words



**1323 Words were removed from one iteration!**  
**The whole dataset can be cleaned given time and computational power**



# Approach



01

## Basic Exploration

Exploring the data

02

## Pre- Processing

Remove null, similar values and stopwords

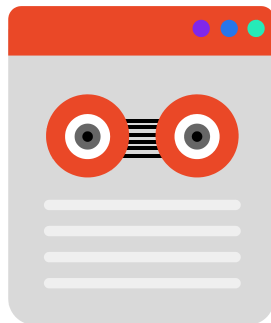
03

## Data Analysis

Data was analysed for checking correctness of various fields

# Future Improvements

**More Data Exploration!**

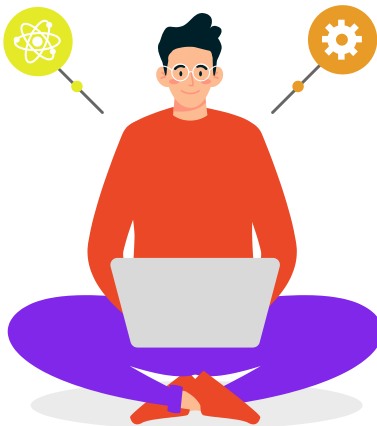


**More Time**

Exam are going on 😓

**New Features**

Working on word2vec



**Yeah That its !**

From all the time i could squeeze in as my exams are going on !

