INNOMATİCS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Fine foods sentiment Analysis using ML

About me

Name : K . VIJAY
Batch : ML Elite 8
Qualification : M.Sc. (Applied Statistics)
Any work experience : Fresher

Name : R.Neharika
Batch : ML Elite 8
Qualification : B.SC(Mathematics)
Any work experience : Fresher

I chose data science because I have a passion for working with data and finding meaningful patterns and insights. It allows me to apply analytical skills to solve real-world problems and make informed decisions. Plus, the field offers exciting career opportunities and the chance to contribute to advancements in various industries

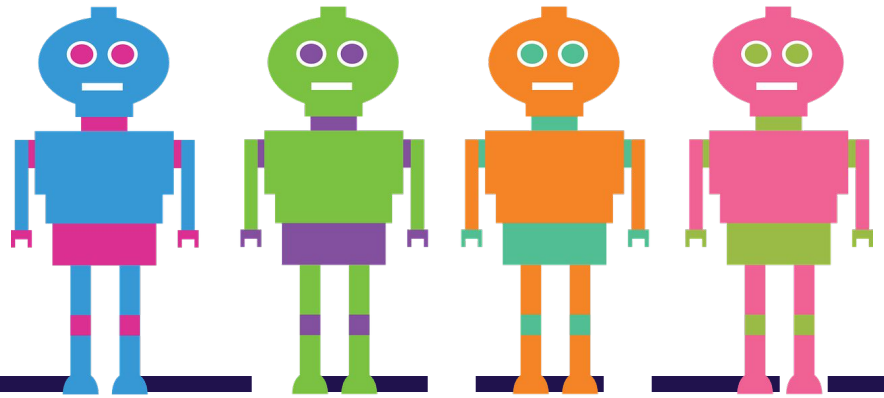
Today's Agenda

- INTRODUCTION
- Clients Expectations
- Process
- Insights
- summary

INNOMATICS RESEARCH LABS

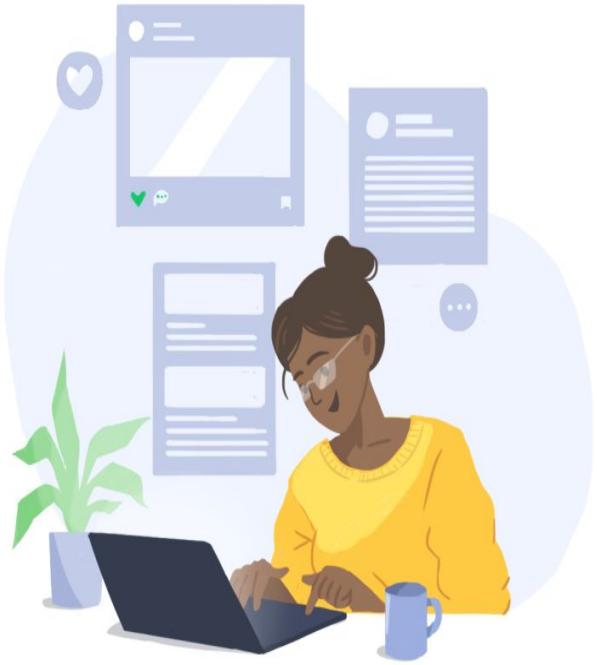
INTRODUCTION

- Dataset consists of reviews of fine foods from amazon
- Reviews from Oct 1999 - Oct 2012
- Reviews include product and user information, ratings, and a plain text review
- Perform data preprocessing on the given text data and convert it into numerical vectors
- Build models to predict the Score of a given text review.



INNOMATICS RESEARCH LABS

Clients Expectations



- To show some nice analysis on given data
- To show the comparison of various ML models

Process

1. Data collection
2. Data cleaning
3. Data analysis
4. Data processing
5. Building models
6. Evaluating models

INNOMATICS RESEARCH LABS

Data cleaning

- Remove any Special characters
- Drop the duplicate column
- Created column helpness using helpness numerator/ helpness denominator
- In column helpness filled null values with 0
- Converted Time column to datetime

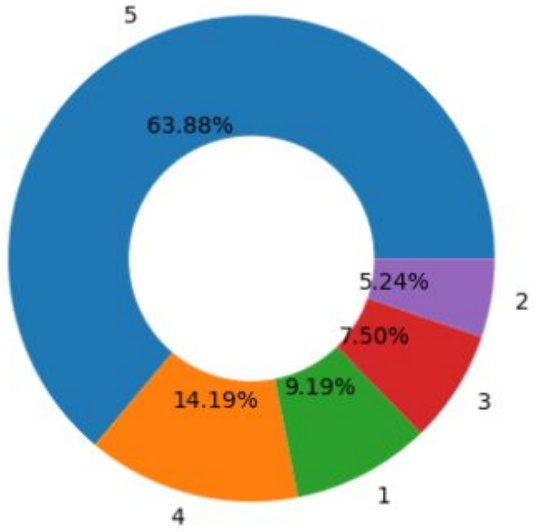
Cleaned Dataframe:

	productid	userid	profilename	helpnessnumerator	helpnessdenominator	score	time	reviewsummary	reviewtext	year
0	B00086IB8O	A3QG1YR5Z5ENYM	Angel D. Rushing	1	1	2	2009-12-26	A little bitter - Not what I expected	I wanted to make a winter wedding cake that h...	2009
1	B004OQADMQ	A1N0RUUU70QRZS	Christiane Shumake	0	0	5	2012-02-06	Not just for babies	I have to give my elderly mother pureed foods...	2012
2	B005T7P2QQ	A235MEPJMXXF66	Frank R. Portscheller	0	0	5	2012-07-15	delicous	I recently came across this very delicous sod...	2012
3	B007RTR8TS	AQCQ5F0QJXMG9	Crystal Starr Light	0	0	5	2012-06-04	Goodbye Garnier Fructus, Hello Clear Scalp & ...	I have thick shoulder-length hair that every ...	2012
4	B003P9XFW8	A1XZU3UA1LDHUM	Kristol King "k2ing1"	0	0	5	2012-10-26	Excellent Dog Food	My dog loves it and she absorbes it well. He...	2012

INNOMATICS RESEARCH LABS

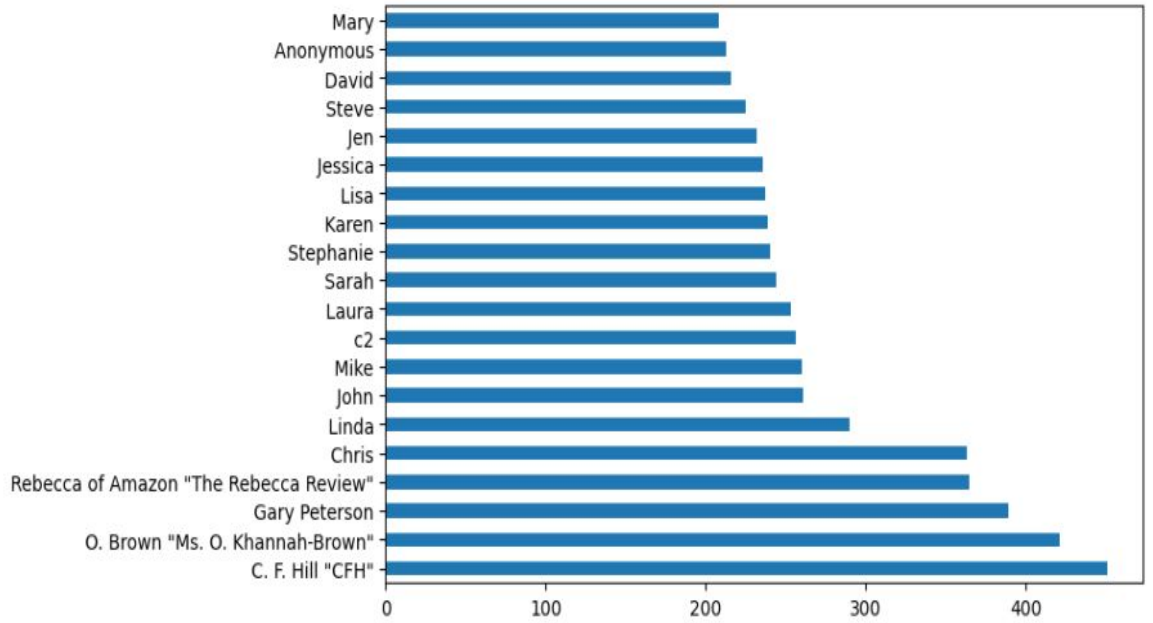
Insights

Score



Maximum number of customers are satisfied and have given 5 score

Profile Names



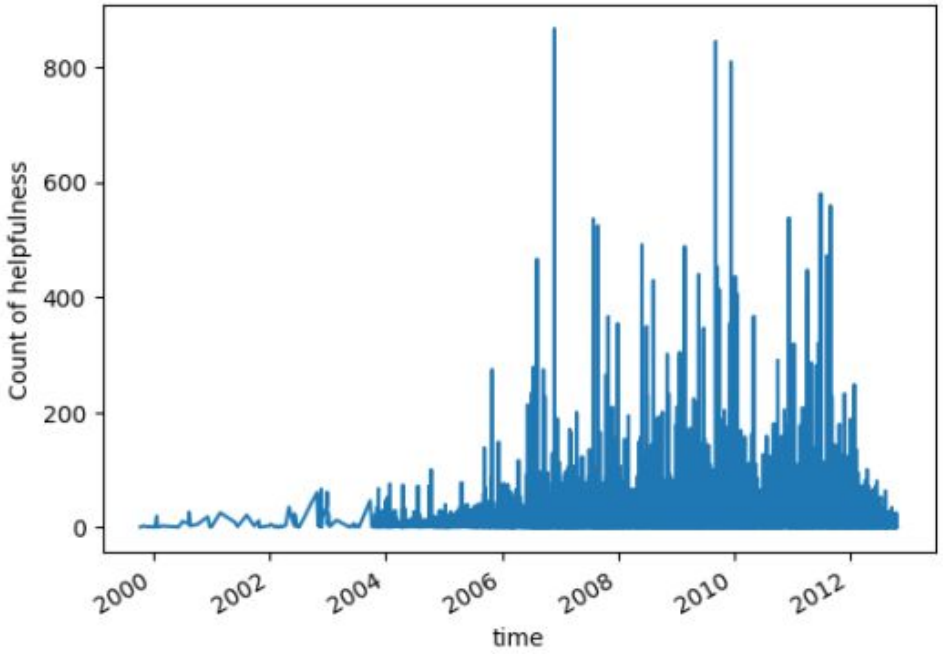
These are top 50 regular customers names

INNOMATICS RESEARCH LABS

Word frequencies

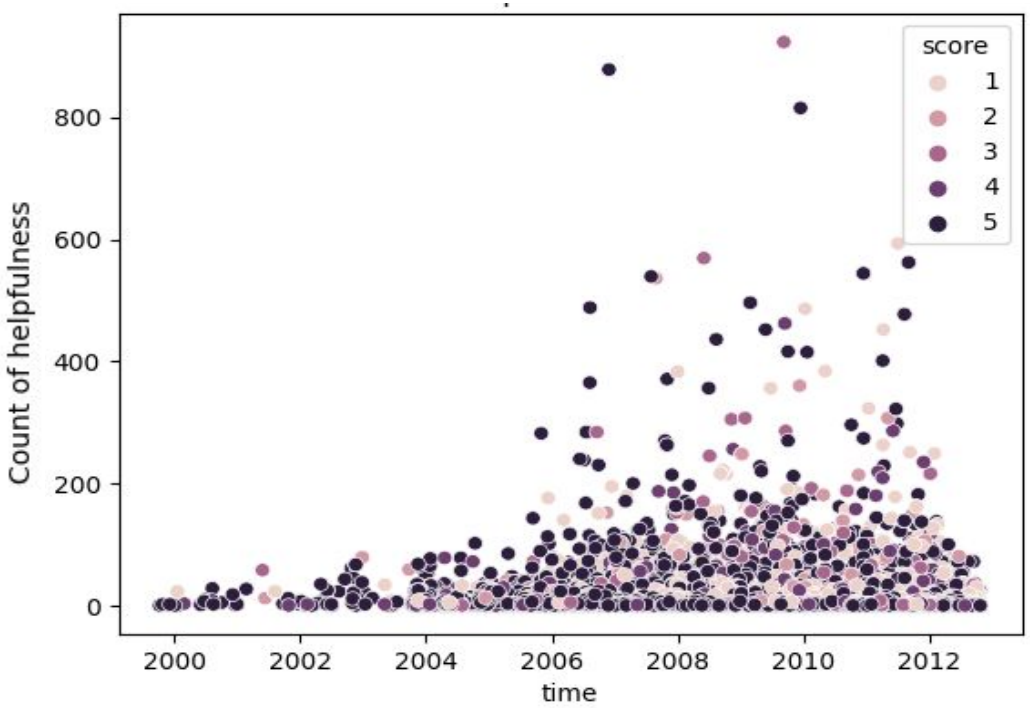
love
good
Best
Delicious
Yummy
taste
Great product
Awesome
better
Perfect
Great Coffee
Great taste
flavor
Wonderful
Excellent
Tasty
gluten free
favorite
Amazing
Fantastic
Pretty good
Good stuff
Good product
coffee
buy
Really
sweet
cup
healthy
Green Tea
great price
Great flavor
Good coffee
Amazon
Yum
Disappointed
Great stuff
dog food
Great tasting
Great Tea
Awesome
dogs love
Perfect
Great Coffee
Great taste
Good stuff
Good product
Pretty good
Wonderful
Yummy
Delicious
favorite
Amazing
Great flavor
Good coffee
Amazon
flavor
Really
buy
coffee
taste
love
good
Best
Tasty
Excellent
Delicious
Yummy
Wonderful
favorite
Amazing
Great flavor
Good coffee
Great taste
Great product
Awesome
better
Perfect
Great Coffee
gluten free
Pretty good
Good stuff
Good product
coffee
buy
Really
flavor

Time vs Helpfulness



With change in time users found review helpful

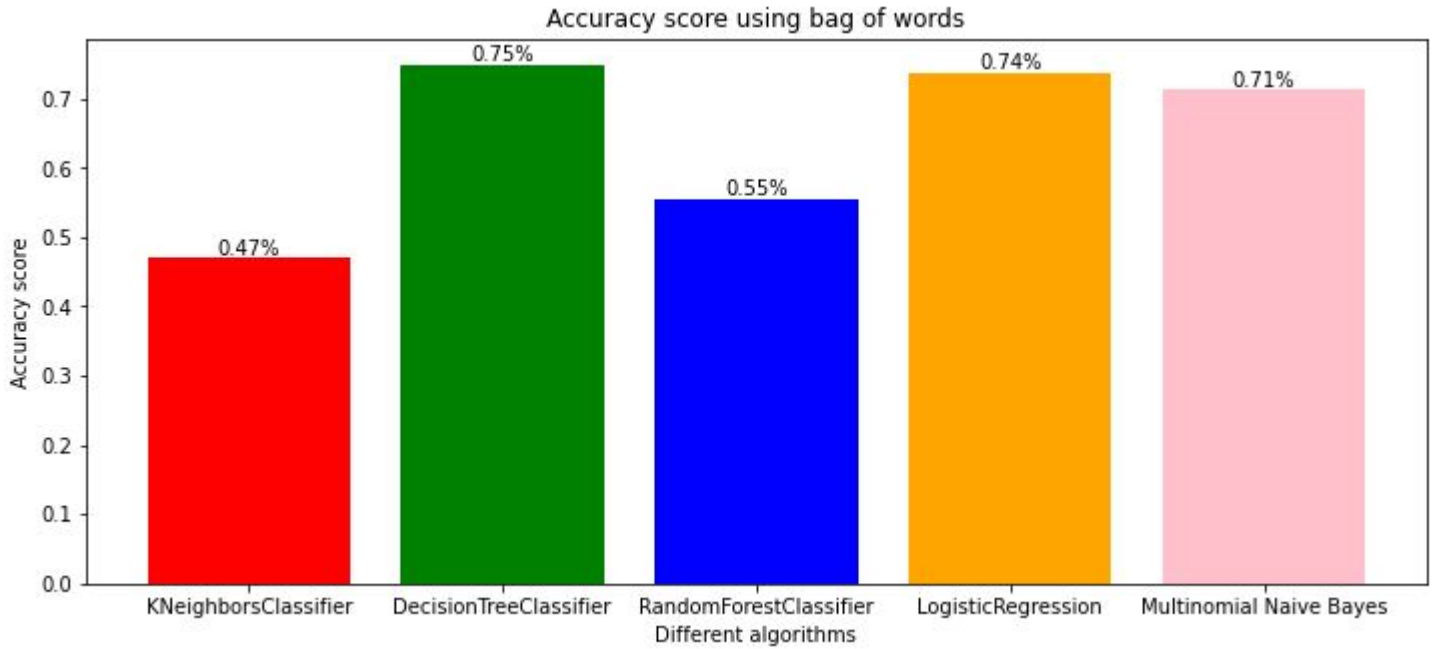
Time vs Helpfulness and score



With time and score users indicated that they found review helpful

INNOMATICS RESEARCH LABS

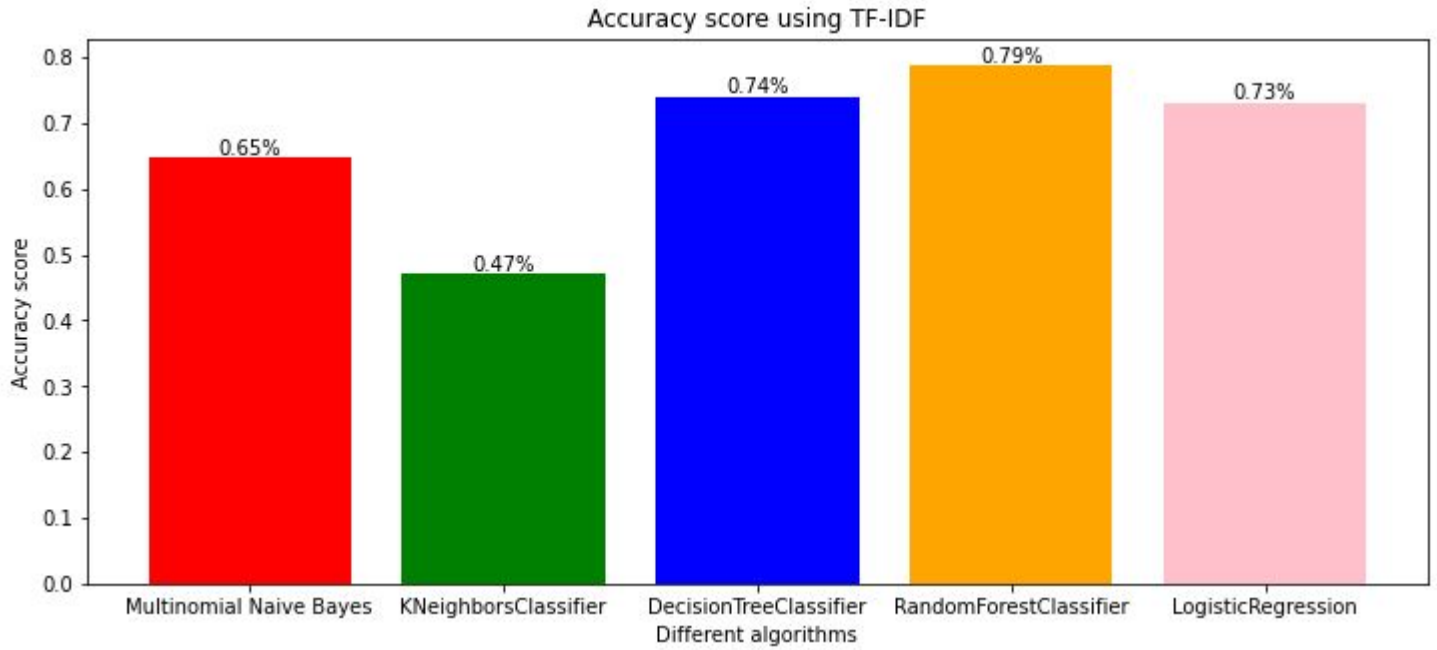
Accuracy Score using Bag of words



- Decision tree classifier performed better when compared to other algorithms using bag of words

INNOMATICS RESEARCH LABS

Accuracy score using TF-IDF



- Random Forest classifier performed better when compared with other algorithms using TF-IDF

INNOMATICS RESEARCH LABS

Observations

- Most of them gave score 5
- With change in time users found review helpful
- Models accuracy using bag of words:

  Decision Tree classifier performed better than other models with 75% accuracy
- Models accuracy using TF-IDF:

  Random Forest classifier performed better than other models with 79% accuracy

INNOMATICS
RESEARCH LABS

INNOMATICS
RESEARCH LABS

THANK YOU

INNOMATICS
RESEARCH LABS