Project Report
on
Comparative Study of Two Food Delivery Giants: Zomato and Swiggy Using Machine Learning and PowerBI

DST-CIMS, INSTITUTE OF SCIENCE

BANARAS HINDU UNIVERSITY

VARANASI-221005

SESSION:2022-2024

Under the Supervision of

Dr. Manjari Gupta

Head of Department

Department of Science and Technology-Centre for Interdisciplinary Mathematical Science

Submitted By:

Anjali Kapoor(22419CAS009)

Project Report
on
Comparative Study of Two Food Delivery Giants: Zomato and Swiggy Using Machine Learning and PowerBI

DST-CIMS, INSTITUTE OF SCIENCE
BANARAS HINDU UNIVERSITY
VARANASI-221005

SESSION: 2022-2024

Under the Supervision of

Dr. Manjari Gupta
Head of Department
Department of Science and Technology-Centre for Interdisciplinary Mathematical Science

Submitted By:
Anjali Kapoor(22419CAS009)

Course Coordinator:
MSc Computational
Science and Applications
DST-CIMS, Institute of Science

Coordinator:
DST- Centre for
Interdisciplinary
Mathematical Science
Institute of Science
BHU, Varanasi-221005

CERTIFICATE

This is to certify that the Dissertation title "**Comparative Study of Two Food Delivery Giants: Zomato and Swiggy Using Machine Learning and PowerBI**", submitted by Anjali Kapoor(22419CAS009)of M.Sc. in Computational Science and Applications (Sem III) during the academic session 2022-2024 in DST-CIMS, Institute of Science, BHU, Varanasi. This work has been carried out under my direct supervision and guidance.

Date:

Dr Manjari Gupta
(Head of Department)
Professor, DST-CIMS
Institute of Science, BHU

ACKNOWLEDGEMENT

I would like to express our sincere gratitude to all those who have contributed to the completion of this thesis. First and foremost, I am deeply thankful to our mentor Dr. Manjari Gupta for her invaluable guidance, support, and unwavering patience throughout the research process and research scholars for their constant support, expertise and insights which have been instrumental in shaping the direction of this study.

My heartfelt thanks go to our Department DST-CIMS of Banaras Hindu University for providing the necessary resources and facilities for the successful completion of this research. I am grateful to our friends and family for their constant encouragement and understanding during the challenging moments of this academic journey. Their emotional support has been a source of strength and motivation. Last but not least, I want to acknowledge the participants of this study whose contributions and cooperation have made this research possible. Thank you all for being a part of this journey and for your invaluable contributions to the completion of this thesis.

Thank you

Anjali Kapoor(22419CAS009)

DECLARATION

I hereby declare that the dissertation titled "Comparative Study of Two Food Delivery Giants: Zomato and Swiggy Using Machine Learning and PowerBI" in partial fulfillment for the award of the Master of Science in Computational Science and Applications is a record of original work carried out by me under the supervision of Dr. Manjari Gupta, Head of Department, DST-CIMS, Institute of Science, BHU. This work has not been submitted for any other degree and professional qualification.

Date:

Name: Anjali Kapoor

TABLE OF CONTENTS

S No.	Content
1.	Abstract
2	Introduction about Zomato and Swiggy
3.	Background and Motivation
4.	Objective
5.	Literature Review
6.	Research Methodology
7.	Discussion
8.	Results and Inferences
9.	Conclusions
10.	Suggestions
11.	Improvements/Future Works
12.	References

Abbreviations

LR	Linear Regression
MAE	Mean Absolute Error
MSE	Mean Squared Error
SVMR	Support Vector Machine Regression
LoR	Logistic Regression
SVMC	Support Vector Machine Classification

1.ABSTRACT

This study aims to conduct a comprehensive comparative analysis of two leading Indian food delivery apps, Swiggy and Zomato, focusing on user satisfaction and app performance. Employing Power BI, we processed data from an extensive survey covering diverse demographic and usage aspects. The objectives included developing a recommendation model for app choice, a prediction model to estimate user ratings based on various parameters, and conducting sentiment analysis on open-ended survey responses.

A total of 227 responses were analyzed to identify key factors influencing user satisfaction and app preference. The study utilized statistical methods to calculate Cronbach's alpha for reliability assessment and constructed a correlation matrix to examine relationships between different variables. Sentiment analysis identified major themes in user feedback, highlighting aspects such as app usability, delivery accuracy, and customer support.

Our findings revealed distinct differences in user experiences and satisfaction levels between Swiggy and Zomato. The recommendation model indicated varying preferences based on factors like app features, frequency of issues encountered, and user engagement. Predictive modeling showed significant predictors of ratings included delivery speed, app interface, and customer service quality. Sentiment analysis demonstrated that users value consistent delivery accuracy and satisfactory issue resolution, while areas such as app reliability and feature variety require improvement.

This comparative study provides actionable insights for both Swiggy and Zomato to enhance their services and establishes a robust methodological approach for analyzing user-generated data in the food delivery sector.

KEYWORDS

Food Delivery Apps, PowerBI, Recommendation Model, Prediction Model, Sentiment Analysis, User Satisfaction, Cronbach's Alpha, Correlation Matrix

2.INTRODUCTION

2.1 About Startup Food Industry

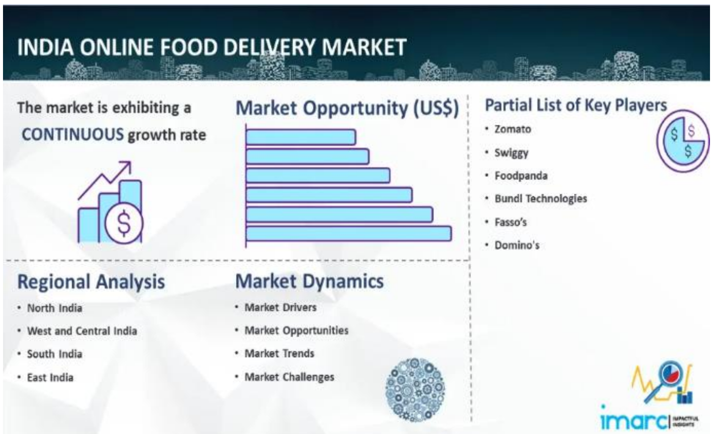


Figure 1

Online food delivery assists individuals in ordering and receiving the desired food products at the doorstep. It involves browsing the website or application, selecting from a wide variety of cuisines available and making the payment through different methods. The website/application updates the user about the expected duration of food preparation and delivery. These features, in confluence with attributes such as ease, speed and precision of delivery, are increasing the demand for these services in India.

The market is currently witnessing growth on account of the increasing access to high-speed internet facilities and the boosting sales of smartphones. This, in confluence with the growing working population and inflating income levels, is propelling the online food delivery market growth in India. Although the players are mainly concentrated in the urban regions of the country, with Bangalore, Delhi and Mumbai representing the three largest markets, vendors are now also targeting smaller cities, as they have strong growth potential. Moreover, the rising trend of on-the-go food items and quick home delivery models that offer convenience, ready-to-eat (RTE) and cheaper food delivery options are escalating the demand for online food delivery services in the country. Furthermore, owing to the rising cases of COVID-19, some of the leading players like Zomato, McDonald's Corporation and Domino's Pizza Inc. have introduced contactless delivery services. These services ensure that the food reaches the customer without being touched by bare hands and is delivered safely with adequate social distancing measures.

Key Market Segmentation:

IMARC Group provides an analysis of the key trends in each sub-segment of the India online food delivery market report, along with forecasts at the country and regional level from 2024-2032. Our report has categorized the market based on platform type, business model and payment method.

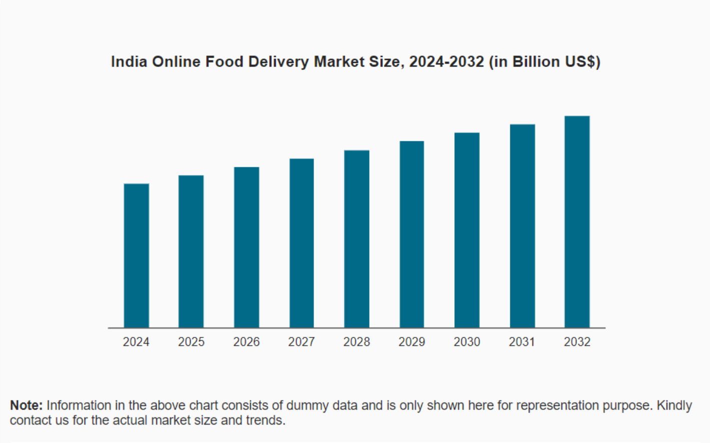


Figure 2

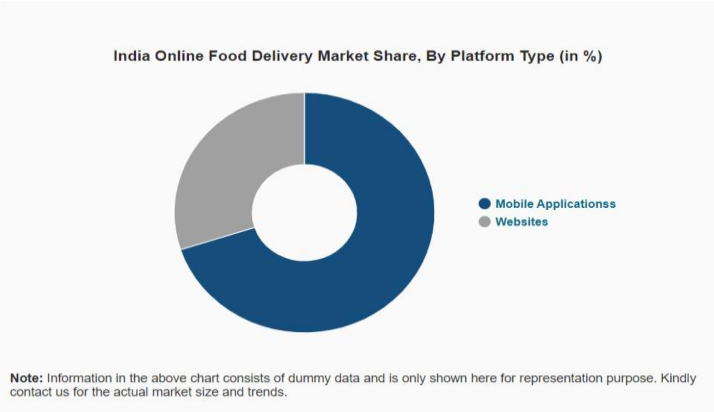


Figure 3

Breakup by Platform Type:

- Mobile Applications
- Websites

Breakup by Business Model:

- Order Focused
- Logistics Based
- Full-Service

Breakup by Payment Method:

- Online
- Cash on Delivery (COD)

Regional Insights:

- North India
- West and Central India
- South India
- East India

Competitive Landscape:

The competitive landscape of the market has been analyzed in the report, along with the detailed profiles of the major players operating in the industry. Some of the leading players include Zomato, Swiggy, Food Panda, Fasso's, Domino's, etc.

2.2 Zomato Company Profile

Figure 4

Zomato(zomato.com) is India's leading restaurant discovery app. It was started by Deepinder Goyal and Pankaj Chaddah in 2008. Zomato works as an aggregator between restaurants and foodies. It helps the foodies to discover, rate, and review the restaurants, and cafes based upon their experience across 10,000 cities in the 24 nations include, India, United States, Australia, United Kingdom, Canada, Turkey, UAE, Qatar, Portugal, South Africa, New Zealand, Chile, Brazil, Indonesia, Philippines, Czech Republic, Poland and Slovakia.

2.3 Zomato History and Origin

Zomato was started with the name 'Foodiebay' in 2008 and later renamed as Zomato in November 2010. It happened because the founders didn't want to confine themselves with only Food Business and also to avoid the confusion with the brand "eBay", because it sounds similar to Foodiebay.

Figure 5-In Frame Pankaj Chaddah and Deepinder Goyal

Firstly, Deepinder Goyal and Pankaj Chaddah started all the menu items from their nearby restaurant and list on their intranet website, after getting some popularity and regular traffic they launched their public website in 2008. After launching the website, they started listing restaurants in Delhi NCR, and quickly extended to Kolkata and then Mumbai.

In 2012, the startup company propelled a print variant of the website in association with Citi bank as "Citibank Zomato Restaurant Guide". Additionally, Zomato extended its presence abroad to the UAE, Sri Lanka, Qatar, UK, Netherlands, Turkey, Brazil, and so on.

Subsequently, Zomato ceaselessly expanded in the worldwide market and peaked in India too. It has a wide reach on its site- about 90 million people visit the site and versatile applications.

Now, Zomato is getting almost 15.2 million traffic per month organically with more than 71 thousand backlinks. (Ahrefs.com)

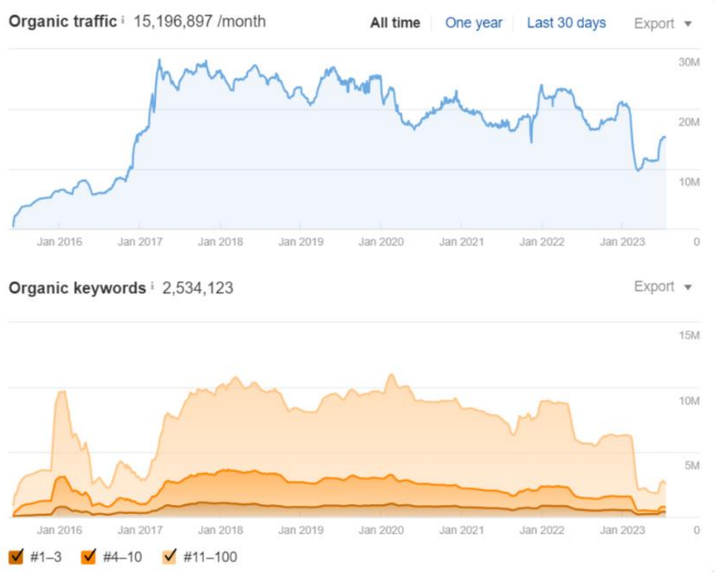


Figure 6-The organic traffic growth chart of Zomato since April 2015.

2.4 Zomato Top Competitor

Though Zomato is very predominantly present in the industry, it does face a lot of direct and indirect competition. Zomato faces direct competition from Swiggy, and competition from other players, including:

- Faasos
- Box8
- Domino's
- FreshMenu
- Pizza Hut
- TravelKhana
- Some of the other international competitors of Zomato are:
- DoorDash
- Uber Eats
- Grubhub Inc.
- Deliveroo
- Postmates
- ChowNow

2.5 Services Offered by Zomato

Some of the prominent products/services of Zomato are:

Zomato Wings: Linking Restaurants and Investors

Zomato unveiled Zomato Wings, a website that links restaurant owners and investors. Serving as a fundraising intermediary, Zomato places a strong emphasis on building a connection between restaurants and venture capital firms to promote expansion in the food sector.

Zomato AI - Revolutionizing Food Discovery

Zomato AI, an innovative AI-powered food discovery companion, is integrated into the platform to redefine how users interact with food-related services. This advanced feature offers personalized suggestions, catering to individual preferences, dietary needs, and moods, revolutionizing the dining experience.

Zomato Future Foundation

Investing in Education: Up to two children of Zomato delivery partners are financially supported by the Zomato Future Foundation, which focuses on education. The project supports employee families and provides further education scholarships for top performance, with an annual coverage of Rs 50,000 per child.

Zomato's Hyperpure

The B2B food tech vertical Hyperpure by Zomato is revolutionizing restaurant operations. With the help of this program, restaurants can purchase premium foods straight from farmers and producers, guaranteeing the consistency, quality, and freshness of their supply.

Zomato Gold

Free deliveries, VIP access during rush hours, and extra savings on dining and delivery services are all included with this exclusive Zomato Gold membership.

Zomaland

Zomato curates an offline carnival called Zomaland that features interactive installations, musicians, comedians, and some of the best restaurants in town. The finest of Zomato Collections are on display at this large event, which provides an immersive experience that goes beyond the screen.

Xtreme

Zomato's parcel delivery app, Xtreme, was released in October 2023 and allows retailers to send and receive tiny parcels. Zomato's revenue streams are diversified and its services are expanded beyond food delivery through Xtreme's utilization of its vast network of delivery partners.

2.6 Zomato Business Model

Zomato not only, earns only a commission from the restaurants, delivery charges, and their membership. That’s not only their overall source of earning money. Zomato is continuously experimenting and growing their business and revenue.

Currently, Zomato earns from in-app advertising, subscription, food delivery service, Zomato Gold. Apart from all these Zomato also organizes Food events and attendees pay a small fee as a ticket.

Online Ordering Service: Zomato takes a certain percentage from the restaurants on each sale done through the Zomato app/website.

In-App Advertising: Advertisement is one of the major contributors to Zomato Revenue. It offers restaurants to buy space or bid on a particular city, on their app and pay for it. It can be a restaurant listing, banners, and sliders. It works similar to Google search and display ads

Delivery Charges: Zomato also charges a delivery fee from the users. It may depend upon your order size and the distance between you and the restaurant. It may vary from Rs.0 to Rs.50.

Organizes Events: Zomato also started organizing food events with a name “Zomaland “. The sale of event tickets and sponsorships contributed to their revenue.

Zomato Book and Zomato Gold: These are 2 services that save a lot of time for the users. With Zomato Book you can reserve any number of seats in any restaurants and cafes with a nominal fee. While Zomato Gold gives some additional benefit to the customer with a monthly or yearly fee like buy 1 get 1 on cold drinks, extra discount, free delivery, and much more. Zomato Gold has been renamed to Zomato Pro.

2.7 Swiggy Company Profile

Swiggy is India's largest online food item ordering & delivery chain, it also tops the chart of India's Unicorn startup lists. It's a Bangalore-based startup started in 2014, and as of now, it's expanded to more than 100 Indian cities. Swiggy propelled quick pick-and-drop food delivery applications to make the life of people simpler. It gives a single window to request from an extensive variety of restaurants along with an entire food entering and conveyance arrangement that connects neighborhood eateries with foodies.

2.8 Swiggy History and Origin

Figure 7

Swiggy came into existence in the year 2014 when two BITS Pilani graduates, Sriharsha Majety and Nandan Reddy came up with the concept 'Hyper-local food delivery'. They got acquainted with Rahul Jaimini, who rejuvenated this vision with a principal site.

In August 2014, Swiggy started activities by joining a couple of eateries in the city of Koramangala in Bengaluru. Following that, they started conveying food to their clients in just 40 minutes.

Soon after this, in May 2015, Swiggy raised its initial round of financing and came up with the application. Through this innovative app, one can get incredible food right to their doorstep and evolve their living standard.

Now, Swiggy is getting almost 5 million traffic per month organically with more than 6.5 thousand backlinks. (Ahrefs.com)

Figure 8-In frame Sriharsha Majety, Nandan Reddy, Rahul Jamini

Figure 9-The organic traffic growth chart of Swiggy since January 2017.

2.9 Swiggy Top Competitor

Swiggy faces direct competition from one very strong rival and that is none other than, of course, Zomato. The latter has a very strong market hold in some cities, for instance, Chandigarh. According to a report in DNA, Swiggy receives a daily order volume of 1.5 million, compared to Zomato’s 1.2 million. After acquiring UberEATS, Zomato has expanded its business to 556 cities and towns. So, the battle between these two food-delivery giants is getting fierce with each passing day.

2.10 Services Offered by Swiggy

As the food ordering and delivery giants, they are now growing more digitally by launching various platforms and increasing their services because of conditions like when you don’t want to go out to the restaurant, then they bring the restaurant to you.

Swiggy has multiple offerings and services for its clients- Swiggy Pop, a single-serve meal delivery service in 30-35 minutes. A selection of items like Indian Thalis, bowl meals, biryanis, burgers, and Asian combos is included in the Swiggy POP menu. All the single-serve meals that come between ranges Rs.99-200 will be delivered in a short time.

The other categories of services that come under Swiggy are Swiggy Cloud, the cloud kitchen service, and Swiggy Stores.

Recently, Swiggy has just launched a platform for partners known as 'Swiggy Partners' where they can request for the premium packaging material. It will help them get the service at a reasonable price and will ensure a better customer experience.

The start-up initially started with delivering basic food items and perishables from nearby stores. In September 2019, it propelled a new service 'Swiggy Go' (very Similar to the Dunzo Business Model) which is used to pick and drop a wide array of items such as clothing, forgotten keys, documents, or deliveries to business persons and other customers. Swiggy Go is currently accessible just in Bangalore, however, said it will expand the service in more than 300 urban cities in a year or so.

2.11 Swiggy Business Model

Swiggy is growing its business at a rapid scale and the revenue generation model of the company is simultaneously expanding.

Delivery Charges: The main sort of income stream Swiggy acquired is from its customers. The company collects delivery charges from customers on the order that costs less than their minimum order of Rs. 250. A charge of 20 to 40 rupees is charged per order.

Commissions: Swiggy acquires another major part of the revenue stream from commissions. It collects commissions from restaurants to generate sales leads and to deliver their food items through Swiggy's application. Restaurants have to pay 15% to 25% on every order placed from Swiggy's site.

Advertisement: Swiggy also procures advertising income in different ways. It shows advertisements of different restaurants on its app and charges address costs to get them promoted in various regions. Also, some restaurants and cafes pay premium rates to Swiggy to get prioritized on the app from the list of accessible eateries.

Swiggy Access: The start-up has come up with the most innovative idea i.e. cloud kitchen idea. It gives its restaurant partners a ready-to-use kitchen area in those zones where they don't work. It brings food closer to its customers and empowers restaurants to set up their kitchen in new areas. Swiggy expects to get 25% of income from the Swiggy access facility and incorporated 30 partners with 36 kitchens.

Swiggy earns subsidiary income as well by collaborating with different financial institutions such as Citibank, HSBC and ICICI Bank. It gives both the parties mutual benefit. It permits clients to get a few charges card offers from these organizations.

3. BACKGROUND AND MOTIVATION

The burgeoning growth of the online food delivery sector in India reflects a paradigm shift in consumer behavior, fueled by technological advancements and changing lifestyles. With platforms like Swiggy and Zomato at the forefront of this transformative wave, understanding the intricacies of their operations, market strategies, and user experiences becomes imperative. As these platforms compete for market share and consumer loyalty, the need for a comprehensive comparative analysis arises to uncover the factors that contribute to their success and distinguish their approaches.

The motivation for this dissertation lies in the recognition of the online food delivery sector as a dynamic and evolving landscape. The increasing reliance on digital platforms for food ordering, combined with the unique strategies employed by Swiggy and Zomato, raises intriguing questions about market dynamics, user preferences, and the overall impact on the culinary ecosystem. The research aims to delve into the nuances of how these platforms have navigated challenges, seized opportunities, and contributed to the reshaping of contemporary dining habits.

Furthermore, this study utilizes Power BI to create interactive and insightful visualizations, transforming raw data into comprehensible charts, graphs, and dashboards. This enables a detailed examination of various metrics and trends, aiding in identifying patterns, making data-driven decisions, and effectively communicating findings to stakeholders. Sentiment analysis on open-ended survey questions using natural language processing techniques helps in understanding user feedback and attitudes. Recommendation models, developed using logistic regression and SVM algorithms, predict user preferences, while prediction models using linear regression and SVM forecast future order behaviors. The performance of these models is compared using metrics such as precision, recall, and F1-score. Additionally, Cronbach's alpha and a correlation matrix are calculated to ensure the reliability and internal consistency of the survey instruments.

In essence, this dissertation is motivated by a broader goal: to contribute to the academic understanding of the digital transformation in the food industry and to provide practical, data-driven insights that can inform strategic decision-making. This research aims to be a valuable resource for industry professionals, policymakers, and researchers interested in the evolving landscape of online food delivery in India.

4. OBJECTIVE

1. To utilize PowerBI for visualizing and analyzing data:

This objective involves employing PowerBI to create interactive and insightful visualizations for the collected data. By transforming raw data into comprehensible charts, graphs, and dashboards, PowerBI enables a detailed examination of various metrics and trends. This visualization aids in identifying patterns, making data-driven decisions, and effectively communicating findings to stakeholders.

2. To perform sentiment analysis on open-ended survey questions:

This objective focuses on analyzing the sentiment expressed in open-ended responses from survey participants. By using natural language processing techniques, the sentiment (positive, negative, or neutral) of the responses will be determined. This analysis helps in understanding the overall sentiment towards specific topics or questions, providing deeper insights into participants' attitudes and opinions.

3. To develop and evaluate recommendation models:

This objective entails building recommendation models using Logistic Regression and Support Vector Machine (SVM) algorithms. These models aim to provide personalized recommendations by analyzing patterns in the data. Feature importance was used to choose the most relevant parameters for these models. Additionally, a correlation matrix will be generated to explore the relationships between different variables. These analyses ensure the robustness of the data collection tools and identify potential interdependencies among variables, enhancing the overall quality of the research. Furthermore, user interfaces (UI) will be created for the app to facilitate the recommendation features, enhancing user interaction and accessibility.

4. To create and evaluate rating prediction models:

This objective involves developing prediction models using Linear Regression and SVM to forecast outcomes based on input data. Feature importance was used to choose the most relevant parameters for these models. The models will be trained to identify relationships between variables and predict rating predictions. Additionally, a correlation matrix will be generated to explore the relationships between different variables. These analyses ensure the robustness of the data collection tools and identify potential interdependencies among variables, enhancing the overall quality of the research. Furthermore, user interfaces (UI) will be designed for the app to support rating prediction features, ensuring a seamless user experience.

5. To compare the accuracy of recommendation and rating prediction models:

This objective focuses on evaluating the performance of the recommendation models and rating prediction models by comparing their accuracy. The models are trained using Logistic Regression, Linear Regression, and SVM algorithms. This comparison provides insights into the strengths and weaknesses of each model, guiding the selection of the most suitable algorithm for the recommendation and rating prediction systems.

6. To calculate Cronbach's alpha and correlation matrix:

This objective aims to assess the reliability and internal consistency of the survey instruments using Cronbach's alpha.

5. LITERATURE REVIEW

Serial No	Title of Paper	Author/Publication	Year of Publication	Findings/Conclusions
1	CONSUMER PERCEPTION TOWARDS ONLINE FOOD ORDERING AND DELIVERY SERVICES: AN EMPIRICAL STUDY	Jyotishman Das (JOM)	2018	Doorstep delivery and ease/convenience drive usage; rewards/cashbacks and location influence decisions. Zomato is preferred. Bad experiences deter use.
2	CONSUMER'S PERCEPTION ON ONLINE FOOD ORDERING	Suryadev Singh Rathore, Mahik Chaudhary (IJMBS)	2018	Youngsters Favor online food ordering, influenced by price, discounts, convenience, and on-time delivery, primarily using Uber Eats or Zomato.
3	UNDERSTANDING CONSUMER BEHAVIOUR TOWARDS UTILIZATION OF ONLINE FOOD DELIVERY PLATFORMS	Chetan Panse, Shailesh Rastogi, Arpita Sharma, Namgay Dorji (IJCSBIE)	2019	Emerging findings: Rising demand for mobile food delivery due to convenience, technology's impact on consumer behavior, and aggregator success during events.
4	A STUDY OF CONSUMER BEHAVIOUR TOWARDS ONLINE FOOD DELIVERY	Rahul Gupta, Sanjoy Roy (Article)	2019	Study focuses on consumer perceptions of food delivery apps, frequency of use, and concerns like food safety during transit. Swiggy rated best.
5	A STUDY ON CONSUMERS PERCEPTION ON FOOD APPS	Aditya Tribhuvan (IJARIIE)	2020	Most people use food apps like Swiggy for convenience and time-saving; cash on delivery is the preferred payment method.
6	A STUDY ON CONSUMER BEHAVIOUR AND THE IMPACT OF FOOD DELIVERY APPS ON THE COLLEGE STUDENTS IN BANGALORE	P. Niharika Nanaiah (IJRESM)	2020	Food delivery app usage is rising, with millennials as key users. Zomato gains edge over Swiggy, which leads the market.

7	CONSUMER BEHAVIOUR AND FACTORS IMPACTING WHILE ORDERING FOOD ONLINE	Bikash Kumar Sahoo, Indronil Sikdar, Dr. Meenal Pendse (IJCRT)	2020	Most respondents are aged 20-25 (93.8%). Pune had the most respondents. Majority spend ₹100-150 on food daily. 78.3% prefer online food ordering. 50% dine out once a week.
8	LITERATURE REVIEW ON CONSUMER PERCEPTION TOWARDS ONLINE FOOD DELIVERY APPS	Gaurav Gawade (Conference: Glocal Evaluation- Through and Post COVID -19 Times)	2021	Consumer behavior towards OFD apps is dynamic, influenced by convenience, internet access, youth usage, product value, discounts, service quality, and timeliness
9	A STUDY ON CONSUMER SATISFACTION ON UBER EATS - AN ONLINE FOOD DELIVERY SYSTEM WITH SPECIAL REFERENCE TO KALAPATTI, COIMBATORE	Gogul Krishna K, Mr.M.A.Prasad, Dr.NGP (IJCRT)	2021	Majority are male (54.5%), aged 20-30 (47.1%), undergraduates (77.4%), students (77.4%), earning Rs. 10001-20000 (45.8%).
10	A STUDY ON CUSTOMERS INFLUENCING FACTORS TOWARDS ONLINE FOOD ORDERING PORTAL SERVICE	Dr. G. Rekha, Dr.V.Santhi (IJCRT)	2021	Convenience is the main factor influencing buying decisions for online food ordering services.
11	REVIEW ON CUSTOMER PERCEPTION TOWARDS ONLINE FOOD DELIVERY SERVICES	Dsouza Prima Frederick, Ganesh Bhat.S (IJCRT)	2021	The study highlights the need for in-depth analysis of factors influencing consumer perception of online food delivery services.

12	UNDERSTANDING THE FACTORS INFLUENCING CUSTOMER'S REVIEWS IN ONLINE FOOD ORDERING SEGMENT	Dr. Amisha Gupta, Rupanshi Tooteja, Bharat Khanna, Rohan Sharma (IJCRT)	2022	Hygiene, taste, delivery time, menu variety, and offers significantly influence online reviews of food businesses.
13	A STUDY ON CUSTOMER PREFERENCE BETWEEN SWIGGY AND ZOMATO: CASE ANALYSIS	Neeraja Pramila, Abhi Patel (IJRPR)	2022	Consumer preferences are vital for brand equity and quality. Online food delivery is popular, emphasizing expediency and ease.
14	A CONSUMER PERCEPTION TOWARDS ONLINE FOOD DELIVERY APPS WITH REFERENCE TO COIMBATORE CITY	Mr. M. Ramesh Kannan, Mrs. S.V. Anitha (IJRPR)	2022	Majority are male (58%), aged 21-30 (51.67%), undergraduates (51.67%), government employees (36%), from Coimbatore (76.67%), earning ₹10,000-20,000 (53%), using Zomato (53%) or Swiggy (50%), ordering weekly (50%), via mobile app (100%).
15	CUSTOMER PERCEPTION TOWARDS ONLINE FOOD DELIVERY SERVICES-DEVELOPMENT OF CONCEPTUAL MODEL	Prima Dsouza, Ganesh Bhat S. (IJCSBIE)	2022	Customers value online food delivery for its convenience and value, but trust and privacy concerns hinder usage. Proposed solutions aim to enhance service experience and attract more users.
16	A STUDY OF CONSUMER BEHAVIOUR TOWARDS FOOD ORDERING AND DELIVERY PLATFORM	Sayali Pachpute (IJIIMS)	2023	70.6% of 18-30-year-olds order food 1-3 times weekly, mainly for dinner; 84.3% of 18-30-year-olds prefer Zomato; 90% use Google Pay, spending over ₹200; 86.2% value convenience; 43.1% of 18-30-year-olds are dissatisfied with loading/delivery times.

17	UNLEASHING THE POWER OF CONVENIENCE: A STUDY OF CONSUMER BEHAVIOR AND MARKET DYNAMICS IN THE ONLINE FOOD DELIVERY INDUSTRY THROUGH THE LENS OF SWIGGY - “SURVIVING AND THRIVING”	Dr Sivaprakash J S, P. Harini (IJCRT)	2023	Respondents prefer Swiggy but cite high delivery fees, slow service, non-eco packaging, and lack detailed menu analysis.
18	EXPLORING CUSTOMER BEHAVIOR IN SWIGGY	A.Hemachandran, Dr. S.Vanithamani, P.M.Sathyanarayanan, S.Syed Moosa Umer (IJCRT)	2023	Swiggy's average order value increased by X%, with 25-34-year-olds ordering most frequently.
19	IMPACT OF COVID-19 PANDEMIC ON ONLINE FOOD DELIVERY ORDERING (OFDO) SERVICES AMONG CONSUMERS SATISFACTION	Parul Sharma (IJCRT)	2023	Customer satisfaction in mobile meal delivery apps is influenced by price, service quality, website quality, and convenience.
20	FACTORS INFLUENCING CONSUMER BEHAVIOR OF SWIGGY & ZOMATO	Kasaram Manasa, Prof. Indrakanti Sekhar (JETIR)	2023	Chi-square: 4629.54, df: 171, p-value: 0.000 (significant), indicating non-identity correlation matrix. KMO: 0.846, suggesting high correlation for factor analysis. PCA reveals Quality and Freshness as key components explaining 67.51% variance.

6. RESEARCH METHODOLOGY

Research is completely based on a logical and systematic way. The study of the overall questions explains with the help of graphs and charts, collecting data from students and analyzing these with logical and scientific tools.

6.1 RESEARCH DESIGN

Type of Methodology: Descriptive Research.

The methodology for research adopted for carrying out the study is:

1st stage – Theoretical/Detailed study was completed.

2nd stage –Customers perception and view towards Zomato and Swiggy.

Descriptive Research:

Descriptive research aims to describe and summarize characteristics of a phenomenon or population, without manipulating variables or attempting to establish causality. It involves observing, measuring, and analyzing data to uncover patterns, trends, and relationships. This type of research provides a snapshot of current conditions or behaviors, answering questions like "what," "who," "where," and "when." Descriptive studies use various methods such as surveys, observations, and archival research to collect data, which is then organized and presented through statistical summaries, charts, and tables. The goal is to provide a comprehensive understanding of the subject being studied, serving as a foundation for further research or decision-making.

6.2 DATASET

Data Collection Method	Non-Probability Method
Technique	Snowball Sampling
Dataset size	227
Area of study	All Over India
Data Analysis Tool	PowerBI, Microsoft Excel, Python

Non-Probability Method:

Non-probability sampling methods involve selecting samples based on subjective judgment or convenience rather than random selection. These methods are commonly used when it's difficult or impractical to obtain a random sample.

It has following types:

1. Convenience Sampling
2. Purposive Sampling
3. Quota Sampling
4. Judgmental Sampling
5. Snowball Sampling

We have used Snowball Sampling for our research:

Snowball Sampling:

Snowball sampling is a non-probability sampling technique where initial participants refer to additional participants, who in turn refer to more participants, creating a chain-like structure. This method is often used to study populations that are hard to reach or hidden, such as marginalized communities or individuals with specific traits or experiences.

In snowball sampling:

1. **Initial Participants**: Researchers start by selecting a few initial participants who meet the criteria for the study. These individuals are typically well-connected within the target population.
2. **Referrals**: Each initial participant is then asked to refer others who also meet the study criteria. These referrals may continue to refer to additional participants, leading to a snowball effect.
3. **Data Collection**: As the sample grows, researchers collect data from each participant through interviews, surveys, or other methods.
4. **Saturation**: The sampling process continues until the desired sample size is reached or until no new information or participants are being added to the study, a point known as saturation.

Snowball sampling is useful when the target population is difficult to access through traditional sampling methods, such as when studying stigmatized groups or individuals with rare characteristics. However, it may introduce biases, as participants tend to refer to others who share similar traits or experiences. Therefore, researchers must carefully consider the limitations and potential biases of snowball sampling when interpreting the results of their studies.

PowerBI:

Power BI is indeed a powerful data visualization tool developed by Microsoft. It allows users to create interactive reports and dashboards from various data sources.

Visualization: Power BI offers a variety of visualization options including bar charts, line charts, scatter plots, maps, tables, and custom visuals. Users can customize the appearance of visuals, apply filters, and create interactive elements for exploration.

Microsoft Excel:

Excel is a versatile tool for data cleaning, it's essential to note that for larger datasets or more complex cleaning tasks, dedicated data cleaning tools or programming languages like Python or R may be more efficient and scalable. Nonetheless, for smaller datasets and quick analyses, Microsoft Excel remains a popular and accessible choice for data cleaning tasks.

Python:

Python is a high-level, interpreted programming language known for its simplicity and readability. It emphasizes code readability and ease of use, making it popular among beginners and professionals alike. Python's extensive standard library and vibrant community support a wide range of applications, from web development to data analysis and machine learning.

6.3 DATA COLLECTION METHOD

There are two major sources of data: Primary Data and Secondary Data

Primary Data- The data which is collected for the first time for a specific purpose. It can be through questionnaires and surveys etc.

Secondary Data- The data which is already available somewhere such as a website, journal etc. We have used both Primary Data. We have prepared a questionnaire in the form of google form and circulated it all over India. Our dataset size is near about 227.

6.4 QUESTIONNAIRE

S. No	QUESTIONS
1	Which Food Delivery app do you use more? Please answer the rest of the questions with respect to your given choice.
2	City
3	State
4	Age of the respondent
5	Gender
6	Educational Qualification
7	Occupation
8	What is your monthly income?
9	If chosen Swiggy, for what purpose do you use the app most?
10	How much do you utilize your chosen App?
11	What type of meals do you most commonly order?
12	What time of day mostly do you place your order?
13	Which mode of payment do you generally use?
14	For whom do you order the most?
15	Have you consistently been using the app, or if there has been a shift in your preference either towards or away from the app over time?
16	If No, why did you shift to/from any other Food delivery App?
17	Preferred cuisines when ordering.
18	On average, how much do you spend per order?
19	Any specific dietary preferences or restrictions?
20	How likely are you to provide a rating after an order?
21	What factors influence your decision when providing a rating?

22	**Do you usually leave written feedback or comments?**
23	**Are you satisfied with the offers given by the app?**
24	**How satisfied are you with the speed of delivery?**
25	**Have you ever experienced issues with the accuracy of the delivery location?**
26	**How would you rate the overall condition of the food upon delivery?**
27	**How satisfied are you with the behavior of the delivery partners?**
28	**Does Swiggy/Zomato recommend top picks according to your past orders?**
29	**How frequently do you encounter issues with the app?**
30	**Which features of the app do you use most often? (Select all that apply)**
31	**Have you ever contacted the customer service?**
32	**How much are you satisfied with the two-way communications with the chat-bot?**
33	**How satisfied were you with the resolution of the issue?**
34	**On a scale of 1 to 10, how would you rate your overall experience with the Swiggy food delivery system? (1 being the lowest, 10 being the highest)**
35	**On a scale of 1 to 10, how would you rate your overall experience with the Zomato food delivery system? (1 being the lowest, 10 being the highest)**
36	**What, if anything, do you think your chosen app does exceptionally well?**
37	**What improvements, if any, would you suggest for your chosen food delivery service?**
38	**Would you prefer your chosen app over other apps?**

6.5 DATA ANALYSIS PROCESS

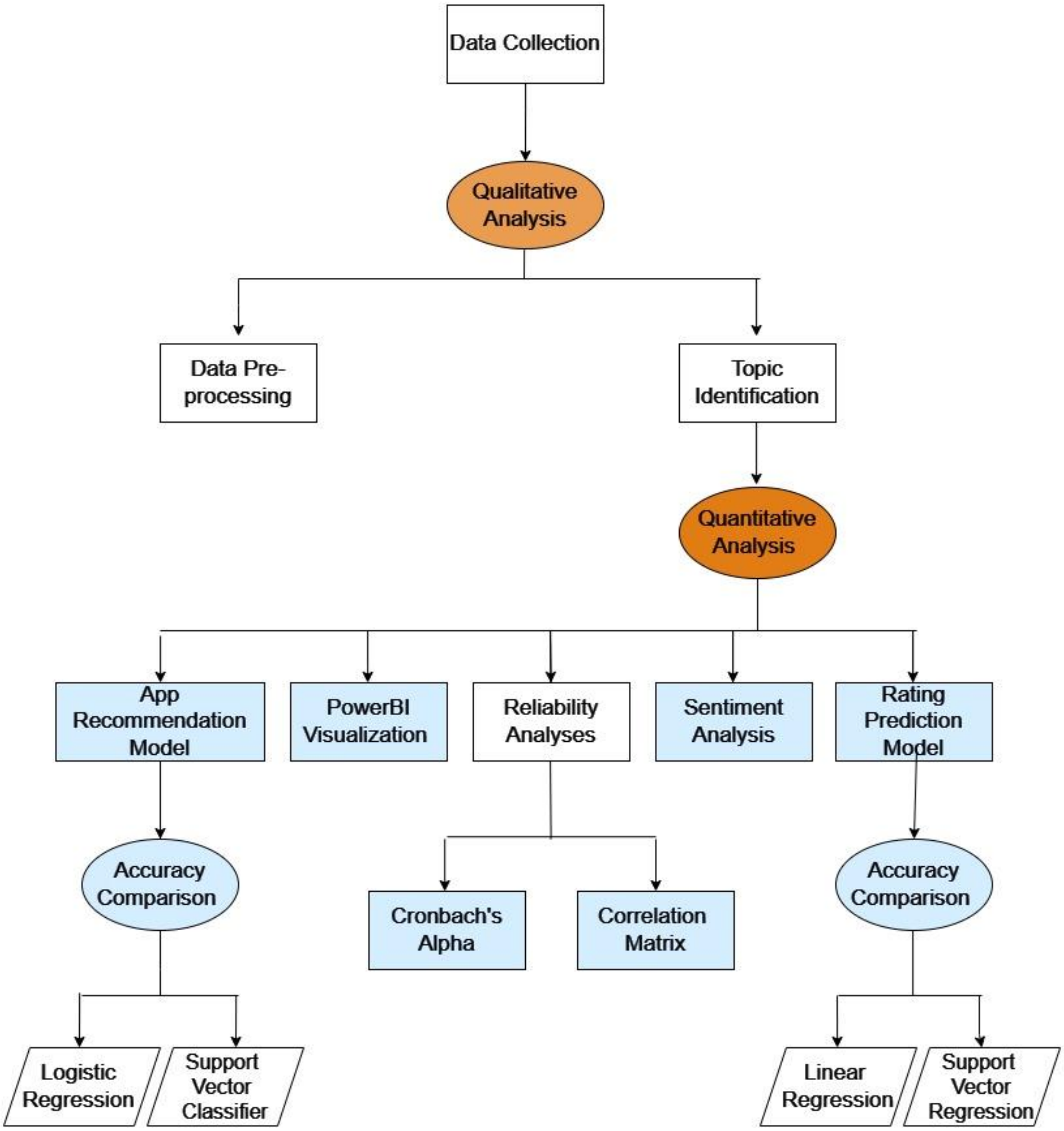


Figure 10

The data analysis process involves the systematic examination, interpretation, and transformation of raw data to derive meaningful insights, draw conclusions, and support decision-making. This process is essential in various fields, including research, business, science, and technology. Here is a general outline of the data analysis process:

- **Data Collection** Firstly we have collected the data through a google form questionnaire.
- **Data Preprocessing** This step involves cleaning and organizing the collected data to prepare it for analysis.
- **Topic Identification** In this stage, the main themes or topics found within the data are identified.
- **Qualitative Analysis** This refers to non-numerical methods used to analyze data.
- **Quantitative Analysis** This refers to numerical methods used to analyze data, such as statistical analysis or machine learning.
- **App Recommendation Model** A machine learning model that recommends apps based on user data.
- **PowerBI Visualization** PowerBI was used to create data visualizations such as charts and graphs.
- **Reliability Analysis** This step refers to assessing the trustworthiness and consistency of the data.
- **Sentiment Analysis** This refers to techniques used to understand the emotional tone of the data, such as positive, negative, or neutral.
- **Rating Prediction Model** A machine learning model that predicts how users might rate an app.
- **Quantitative Analysis** This branch shows several data analysis techniques including:
    - Correlation Matrix: This shows the correlation between different variables in the data set.
    - We have used two models Logistic Regression and Support Vector Classifier to predict the given app on the basis of some parameters.
    - We have used two models Linear Regression and Support Vector Regression to predict the rating of the app on the basis of some parameters.
- **Accuracy Comparison** This could refer to comparing the accuracy of the App Recommendation Model and the Rating Prediction Model.

7. DISCUSSIONS

7.1 Cronbach's Alpha

Cronbach's alpha is a measure of internal consistency or reliability of a set of scale or test items. It is commonly used to assess the reliability of a survey or questionnaire that consists of multiple Likert-type questions.

What is Cronbach's Alpha?

Cronbach's alpha (α) is a statistic that indicates how well a set of items measures a single unidimensional latent construct. It ranges from 0 to 1, with higher values indicating greater internal consistency.

Assumptions

1. One-dimensionality: The items should measure a single construct.
2. Tau-equivalence: The items are assumed to have the same true score variance.
3. Normally Distributed Errors: The errors associated with each item should be normally distributed.

Limitations

1. One-dimensionality: Cronbach's alpha assumes that the scale measures a single construct. If the scale is multidimensional, alpha may underestimate the reliability.
2. Tau-equivalence: Assumes equal true score variance for each item, which is often not the case in practice.
3. Not a Measure of One-dimensionality: High alpha does not imply that the scale is unidimensional.
4. Affected by the Number of Items: Alpha tends to increase with the number of items, which might give a false sense of reliability.

$$\alpha = \frac{\overset{\text{Number of items}}{N}\bar{c}}{\underset{\text{Average variance}}{\bar{v}} + (N-1)\underset{\text{Average inter-item covariance among the items}}{\bar{c}}}$$

Figure 11

7.2 Emotional Analytics: Extracting Sentiment from Textual Data

What is Sentiment Analysis?

Figure 12

Sentiment analysis is a process of computationally identifying and categorizing opinions expressed in text data, determining whether the sentiment conveyed is positive, negative, or neutral. It involves natural language processing techniques to analyze the tone, emotions, and subjective information present in the text. By examining words, phrases, and context, sentiment analysis algorithms can extract valuable insights from large volumes of text, enabling businesses to understand public opinion, customer feedback, and market trends. Sentiment analysis provides a quantitative measure of sentiment, aiding decision-making, reputation management, and customer relationship management efforts.

Compound Value

Sentiment analysis scores using compound values typically involve assessing the overall sentiment polarity of a piece of text by considering various linguistic cues, such as words, phrases, and their respective intensities.

The compound value represents a comprehensive sentiment score ranging from -1 (indicating extremely negative sentiment) to 1 (indicating extremely positive sentiment), with 0 suggesting a neutral sentiment.

Why is Sentiment Analysis Important?

Since humans express their thoughts and feelings more openly than ever before, sentiment analysis is fast becoming an essential tool to monitor and understand sentiment in all types of data.

Automatically analysing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs.

Sorting Data at Scale

Can you imagine manually sorting through thousands of tweets, customer support conversations, or surveys? There's just too much business data to process manually. Sentiment analysis helps businesses process huge amounts of unstructured data in an efficient and cost-effective way.

Real-Time Analysis

Sentiment analysis can identify critical issues in real-time, for example is a PR crisis on social media escalating? Is an angry customer about to churn? Sentiment analysis models can help you immediately identify these kinds of situations, so you can take action right away.

Consistent Criteria

It's estimated that people only agree around 60-65% of the time when determining the sentiment of a particular text. Tagging text by sentiment is highly subjective, influenced by personal experiences, thoughts, and beliefs.

By using a centralized sentiment analysis system, companies can apply the same criteria to all of their data, helping them improve accuracy and gain better insights.

How does Sentiment Analysis work?

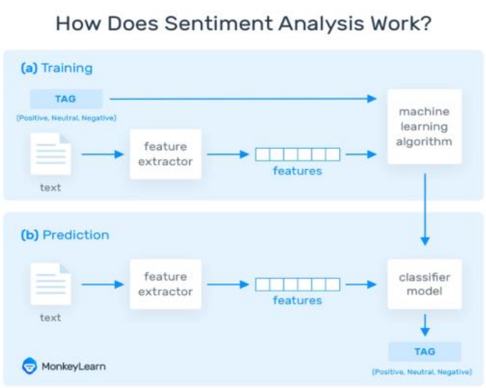


Figure 13

Sentiment analysis, otherwise known as opinion mining, works thanks to natural language processing (NLP) and machine learning algorithms, to automatically determine the emotional tone behind online conversations.

There are different algorithms you can implement in sentiment analysis models, depending on how much data you need to analyse, and how accurate you need your model to be. We'll go over some of these in more detail, below.

Sentiment analysis algorithms fall into one of three buckets:

Rule-based: These systems automatically perform sentiment analysis based on a set of manually crafted rules.

Automatic: Systems rely on machine learning techniques to learn from data.

Hybrid systems: Combine both rule-based and automatic approaches.

Rule-based Approaches

Usually, a rule-based system uses a set of human-crafted rules to help identify subjectivity, polarity, or the subject of an opinion.

These rules may include various NLP techniques developed in computational linguistics, such as: Stemming, tokenization, part-of-speech tagging and parsing. Here's a basic example of how a rule-based system works:

Defines two lists of polarized words (e.g. negative words such as bad, worst, ugly, etc and positive words such as good, best, beautiful, etc).

If the number of positive word appearances is greater than the number of negative word appearances, the system returns a positive sentiment, and vice versa. If the numbers are even, the system will return a neutral sentiment.

Rule-based systems are very naive since they don't take into account how words are combined in a sequence. However, adding new rules may affect previous results, and the whole system can get very complex. Since rule-based systems often require fine-tuning and maintenance, they'll also need regular investments.

Automatic Approaches

Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. A sentiment analysis task is usually modelled as a classification problem, whereby a classifier is fed a text and returns a category, e.g. positive, negative, or neutral.

Hybrid Approaches

Hybrid systems combine the desirable elements of rule-based and automatic techniques into one system. One huge benefit of these systems is that results are often more accurate.

7.3 Word Cloud

A word cloud generated through sentiment analysis visually represents the frequency of words in a text corpus, with the size of each word proportional to its occurrence. However, unlike a traditional word cloud that merely displays word frequency, a sentiment analysis word cloud incorporates sentiment information to highlight the emotional tone associated with each word.

Word Frequency: Like in a traditional word cloud, the size of each word corresponds to its frequency in the text corpus. Commonly occurring words appear larger, while less frequent ones appear smaller.

Emotion Visualization: By incorporating sentiment analysis, the word cloud offers a visual depiction of the emotional content within the text corpus. Words associated with strong emotions stand out, providing insights into the prevailing sentiments expressed in the analysed text.

Contextual Understanding: In some cases, sentiment analysis word clouds may include contextual information such as word groupings or associations to convey more nuanced insights into sentiment patterns within the text.

Overall, a sentiment analysis word cloud provides a visually engaging and informative representation of sentiment-laden words within a text corpus, aiding in the interpretation and exploration of emotional content across various domains such as customer feedback, social media conversations, or product reviews.

What is Sentiment Polarity?

Sentiment polarity refers to the direction or orientation of sentiment expressed in a piece of text, indicating whether the sentiment conveyed is positive, negative, or neutral.

In sentiment analysis, polarity is typically quantified on a numerical scale, with values ranging from -1 to 1:

- A polarity score of -1 represents extremely negative sentiment.
- A polarity score of 1 represents extremely positive sentiment.
- A polarity score of 0 indicates neutral sentiment, meaning there is an absence of strong positive or negative sentiment.

Sentiment polarity analysis involves analysing the text's language and context to determine the overall emotional sentiment conveyed. This analysis often involves natural language processing (NLP) techniques, such as text classification algorithms, to classify the sentiment of the text accurately.

7.4 Foodie's Dilemma: Zomato or Swiggy? Recommending the Best Food Delivery App

What is a Recommendation system?

A recommendation system is a type of machine learning system that provides personalized recommendations to users based on their past behaviours, preferences, and patterns. It is a subclass of information filtering systems that use algorithms to recommend items to users based on their interests or behaviours.

Figure 14

Recommendation systems are widely used in e-commerce, social media, entertainment, and other online platforms to increase user engagement and retention, improve customer satisfaction, and drive sales and revenue.

How do recommendation system work?

Here are the 4 steps of how recommendation systems work:

1. **Collecting user data:** The first step in building a recommendation system is to collect user data. This can include user ratings, reviews, clickstream data, purchase history, and other behavioural data. The data can be collected either explicitly, through user surveys or feedback forms, or implicitly, through user interactions with the platform.
2. **Storing the data:** Once the user data is collected, it needs to be stored in a database or data warehouse for analysis. The data can be stored in a structured or unstructured format, depending on the type and volume of the data.
3. **Analysing the data:** The next step is to analyse the user data to identify patterns and trends. This can be done using various data analysis techniques like clustering, classification, and regression. The goal is to understand the user’s preferences, behaviours, and interests, and to use this information to make personalized recommendations.
4. **Filtering and recommending**: The final step is to filter the data and make recommendations to the user. This can be done using various recommendation algorithms, such as collaborative, content-based, and hybrid filtering. The algorithm uses the user data and the analysis results to generate a list of recommended items the

user will likely be interested in. The recommendations are then presented to the user in a personalized way, such as through a recommendation widget, email, or push notification.

Now, there are three main types of recommendation systems:

1. Content-Based Filtering:

Content-based recommendation systems recommend items to users based on their past preferences and behaviours. This type of system analyses the user's historical data, such as their search history, browsing history, or purchase history, and recommends items that are like the ones the user has interacted with before.

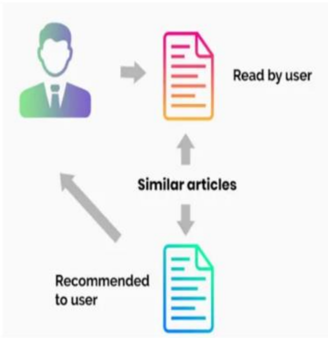


Figure 15

2. Collaborative Filtering

Collaborative filtering recommendation systems recommend items to users based on the preferences and behaviours of other similar users. This type of system analyses the user's historical data, as well as the data of other users with similar preferences, and recommends items that similar users have liked or interacted with before.

There are two kinds of collaborative filtering techniques:

Figure 16

- User-User collaborative filtering
- Item-Item collaborative filtering

User-User collaborative filtering is a type of recommendation system that makes predictions for a user based on the preferences of similar users. It works by finding users with similar tastes and recommending items they liked to the target user. Item-Item collaborative filtering, on the other hand, recommends items to a user based on the preferences for similar items. It works by identifying items that are like the ones a user has liked in the past and recommending them to the user.

3. Hybrid Recommendation Systems

Hybrid recommendation systems combine both content-based and collaborative filtering techniques to provide more accurate and diverse recommendations. This type of system uses a combination of user data, item data, and other contextual information to generate recommendations.

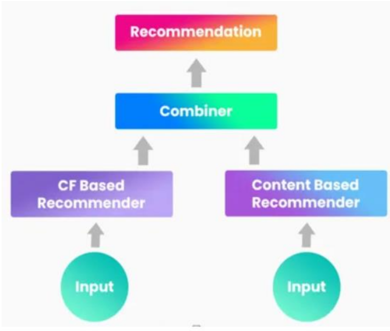


Figure 17

For our research, we have used **User-User Collaborative Filtering** because our recommendation model will **recommend the best food delivery app** according to the previous user's preferences to the new user/target user who has the similar taste.

7.5 Comparative Rating prediction: Zomato v/s Swiggy

What is a Prediction Model?

Prediction models are built using machine learning algorithms or statistical techniques that analyse historical data to identify patterns and relationships. These models can be used to predict a wide range of outcomes, such as sales figures, customer behaviour, disease progression, stock prices, or weather patterns. To create a prediction model, data scientists typically follow a series of steps**,** including data collection, data preprocessing, feature selection, model training, model evaluation, and deployment. The model is trained on a portion of the historical data and then tested on another portion to assess its accuracy and performance.

Figure 18

There are various types of prediction models, including linear regression, decision trees, neural networks, support vector machines, and random forests, each with its own strengths and limitations. The choice of model depends on the nature of the data and the specific prediction task at hand.

Top five predictive analytics models are:

Classification model: Considered the simplest model, it categorizes data for simple and direct query response. An example use case would be to answer the question "Is this a fraudulent transaction?"

Clustering model: This model nests data together by common attributes. It works by grouping things or people with shared characteristics or behaviours and plans strategies for each group at a larger scale. An example is in determining credit risk for a loan application.

An example is in determining credit risk for a loan applicant based on what other people in the same or a similar situation did in the past.

Forecast model: This is a very popular model, and it works on anything with a numerical value based on learning from historical data. For example, in answering how much lettuce a restaurant should order next week or how many calls a customer support agent should be able to handle per day or week, the system looks back to historical data.

Outliers model: This model works by analysing abnormal or outlying data points. For example, a bank might use an outlier model to identify fraud by asking whether a transaction is outside of the customer's normal buying habits or whether an expense in each category is normal or not. For example, a $1,000 credit card charge for a washer and dryer in the cardholder's preferred big box store would not be alarming, but $1,000 spent on designer clothing in a location where the customer has never charged other items might be indicative of a breached account.

Time series model: This model evaluates a sequence of data points based on time. For example, the number of stroke patients admitted to the hospital in the last four months is used to predict how many patients the hospital might expect to admit next week, next month or the rest of the year. A single metric measured and compared over time is thus more meaningful than a simple average.

For our research, we have used the **Forecast Model** wherein we **predict the overall rating of Zomato/Swiggy** using some relevant features.

7.5.1 Determining Key Factors: The Role of Feature Importance

Feature importance refers to a technique used in machine learning to determine the significance or contribution of each feature (input variable) to the predictive power or performance of a model. It helps to identify which features have the most impact on the outcome variable and can guide decisions regarding feature selection, model interpretability, and understanding underlying data patterns.

Feature importance is a crucial aspect of model interpretability and performance optimization. By understanding which features are most influential, we can improve model accuracy, reduce complexity, and gain insights into the data.

Here, we have used **Random Forest Feature Importance Method** for determining importance of each feature.

The Random Forest method for determining feature importance is based on the ensemble learning technique that constructs multiple decision trees during training. In this method, the importance of a feature is evaluated by observing how much it contributes to reducing impurity, which is typically measured using criteria like Gini impurity or entropy.

7.5.2 Correlation Matrix: Identifying Key Parameter Relationships

A correlation matrix is a table showing correlation coefficients between sets of variables. Each cell in the table shows the correlation between two variables. This matrix is used to summarize data, as an input for advanced analyses, and as a diagnostic tool.

The values range from -1 to 1.

- 1: Perfect positive correlation. As one variable increases, the other also increases proportionally.
- -1: Perfect negative correlation. As one variable increases, the other decreases proportionally.
- 0: No correlation. The variables do not have any linear relationship.

7.6 Assessing Model Reliability: Accuracy Comparison on Swiggy and Zomato Data

What is Machine Learning?

Machine learning is a subset of artificial intelligence that involves the development of algorithms and statistical models that enable computer systems to learn from and make predictions or decisions based on data without being explicitly programmed. In other words, machine learning algorithms use patterns and insights from data to improve their performance over time without human intervention.

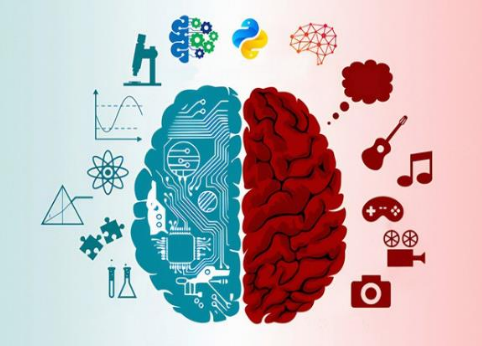


Figure 19

Types of Machine Learning Models:

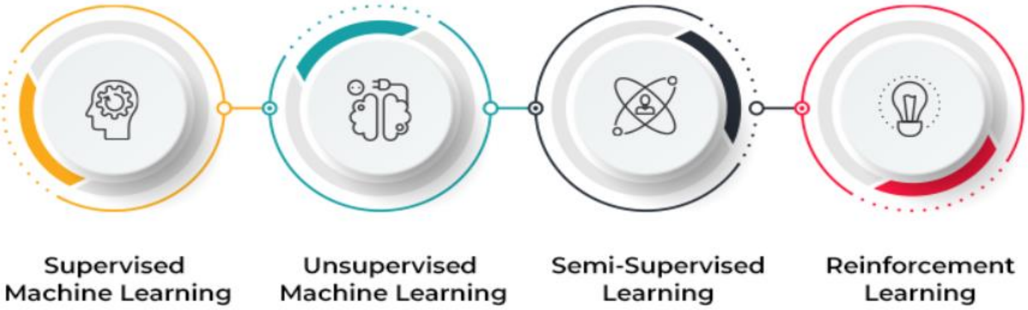


Figure 20

Supervised Learning:

Supervised learning is a type of machine learning where the model learns to map input data to output labels based on example input-output pairs provided in a training dataset. The term "supervised" refers to the process of supervising or guiding the model's learning process by providing labelled data for training.

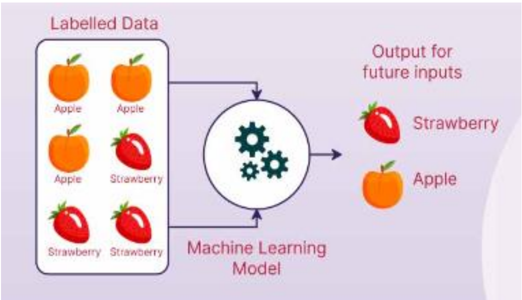


Figure 21

Key Components:

1. **Input Data (Features):**

   Input data, also known as features or predictors, are the variables or attributes used to make predictions. These can be numerical, categorical, or even text data, depending on the problem domain.

2. **Output Labels (Target Variable):**

   Output labels, also known as the target variable or response variable, are the variables the model aims to predict. In regression tasks, the target variable is continuous, while in classification tasks, it is categorical.

3. **Training Data:**

   The training data consists of a set of example input-output pairs used to train the supervised learning model. Each example includes input features and their corresponding output labels.

4. **Model:**

   The model is a mathematical function that maps input features to output labels. During the training process, the model learns from the training data to make accurate predictions on unseen data.

5. **Loss Function:**

   A loss function quantifies the difference between the model's predictions and the true output labels in the training data. The goal during training is to minimize this loss function, effectively improving the model's predictive performance.

6. **Evaluation:**

   Once trained, the supervised learning model needs to be evaluated on a separate dataset called the test data to assess its performance. Common evaluation metrics for regression tasks include mean squared error (MSE) or mean absolute error (MAE), while for classification tasks, metrics such as accuracy, precision, recall, and F1-score are often used.

   In summary, supervised learning is a fundamental technique in machine learning, where models are trained to learn patterns from labelled data to make predictions on unseen data. It forms the basis for many real-world applications and continues to be an active area of research and development.

Unsupervised Learning:

Unsupervised learning is a branch of machine learning where the model learns patterns and structures from input data without explicit supervision or labelled examples. Unlike supervised learning, where the algorithm is trained on labelled data, unsupervised learning algorithms infer patterns and relationships directly from the input data itself.

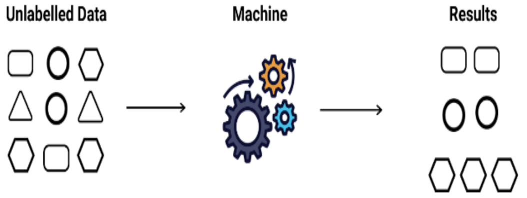


Figure 22

Key Concepts:

1. **Input Data:**

   Unsupervised learning algorithms operate on input data without any corresponding output labels. This input data could be raw features, images, text, or any other type of data.

2. **Objective:**

   The primary objective of unsupervised learning is to uncover hidden patterns, structures, or relationships within the data. This can include identifying clusters of similar data points, reducing the dimensionality of the data, or discovering underlying trends.

3. **No Ground Truth:**

   Unlike supervised learning, where there is a ground truth provided by labelled data, unsupervised learning does not have access to such information. Instead, the model must infer patterns solely from the input data.

4. **Clustering:**

   Clustering is one of the main tasks in unsupervised learning, where the goal is to partition the input data into groups or clusters of similar data points. Common clustering algorithms include k-means, hierarchical clustering, and DBSCAN.

5. **Dimensionality Reduction:**

   Dimensionality reduction techniques aim to reduce the number of features in the data while preserving as much relevant information as possible. Principal Component Analysis (PCA) and t-distributed Stochastic Neighbour Embedding (t-SNE) are popular dimensionality reduction algorithms.

6. **Anomaly Detection:**

   Anomaly detection involves identifying unusual or rare data points that deviate from the norm. Unsupervised learning techniques can be used to detect anomalies in various applications such as fraud detection, network security, and predictive maintenance.

7. Evaluation:

Evaluating the performance of unsupervised learning algorithms can be challenging since there is no ground truth to compare the results against. Evaluation typically involves qualitative assessment by domain experts or using metrics specific to the task, such as silhouette score for clustering or reconstruction error for dimensionality reduction.

In summary, unsupervised learning is a powerful approach for exploring and discovering patterns in data without the need for labelled examples. It plays a crucial role in understanding complex datasets and extracting valuable insights that can drive decision-making and innovation in diverse fields.

Semi - Supervised Learning:

Semi-supervised learning is a machine learning paradigm that falls between supervised and unsupervised learning. It combines the use of both labelled and unlabelled data to improve the performance of predictive models. Here's a detailed explanation of semi-supervised learning:

Key Concepts:

1. Labelled and Unlabelled Data:

In semi-supervised learning, the dataset contains a mixture of labelled data, where each example is associated with a target label, and unlabelled data, where the target labels are not provided.

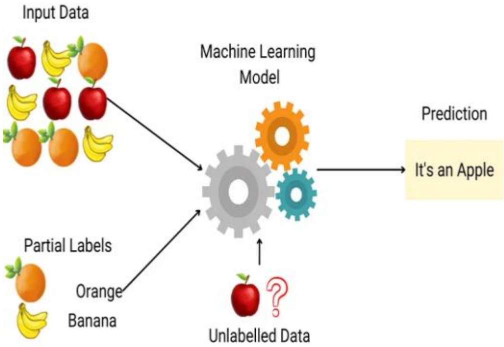


Figure 23

2. Limited Availability of Labelled Data:

In many real-world scenarios, obtaining labelled data can be expensive, time-consuming, or even impractical. Semi-supervised learning addresses this challenge by leveraging the abundance of unlabelled data and a small amount of labelled data.

3. Utilizing Unlabelled Data:

Semi-supervised learning algorithms use the information present in the unlabelled data to improve the model's performance. By learning from the underlying structure or distribution of the data, the model can generalize better to unseen examples.

4. **Pseudo-labelling:**

One common approach in semi-supervised learning is pseudo-labelling, where the model generates labels for the unlabelled data based on its predictions. These pseudo-labels are then combined with the labelled data to train the model in a supervised manner.

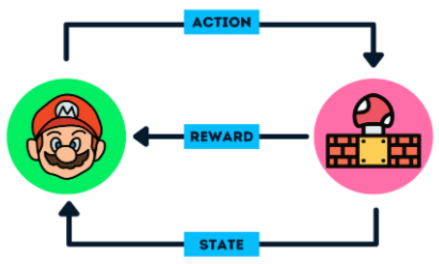


Figure 24

5. **Combining Supervised and Unsupervised Techniques:**

Semi-supervised learning algorithms often combine techniques from both supervised and unsupervised learning. For example, they may use clustering or dimensionality reduction methods to exploit the structure of the unlabelled data, in addition to traditional supervised learning techniques.

6. **Active Learning:**

Active learning is another approach used in semi-supervised learning, where the model selects the most informative examples from the unlabelled data for manual labelling. By focusing on the most uncertain or ambiguous examples, the model can learn more effectively with limited labelled data.

In summary, semi-supervised learning offers a practical and efficient approach to leveraging both labelled and unlabelled data, enabling improved model performance and generalization in various real-world applications.

Reinforcement Learning:

Reinforcement learning (RL) is a type of machine learning paradigm where an agent learns to make decisions by interacting with an environment to achieve a goal. Unlike supervised learning, where the model learns from labelled input-output pairs, and unsupervised learning, where the model discovers patterns in unlabelled data, reinforcement learning learns through trial and error by receiving feedback from the environment. Here's a comprehensive explanation of reinforcement learning:

Key Concepts:

1. **Agent:**

   The agent is the entity that interacts with the environment in reinforcement learning. It makes decisions based on its observations of the environment and receives feedback in the form of rewards or penalties.

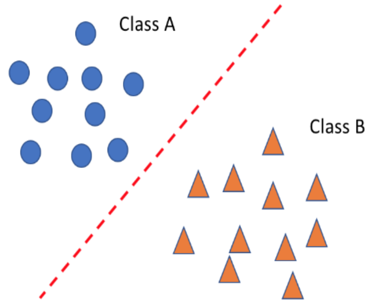


Figure 25

2. **Environment:**

   The environment is the external system with which the agent interacts. It can be as simple as a grid world or as complex as a simulated physical environment. The environment provides feedback to the agent based on its actions.

3. **Actions:**

   Actions are the decisions made by the agent to affect the state of the environment. The agent selects actions based on its current state and a policy, which defines the agent's behaviour.

In our research, we have used Support Vector Machine Classification and Logistic Regression **for classifying between the two apps and then recommending the most suitable app to the user**.

Support Vector Machine Classification

Support Vector Machine (SVM) is a supervised learning algorithm used for classification and regression tasks. It is particularly effective in high-dimensional spaces and is widely used in various fields such as image recognition, bioinformatics, and text classification.

Here's a comprehensive overview of the theory behind Support Vector Machines:

1. **Basics of SVM:**

   SVM is a discriminative classifier that separates data points into different classes using a hyperplane.

   It works by finding the optimal hyperplane that maximizes the margin, the distance between the hyperplane and the nearest data points of each class, known as support vectors.

2. **Hyperplane:**

   In a binary classification problem with two classes, a hyperplane is a decision boundary that separates the data points into two classes.

   In an n-dimensional space, a hyperplane is represented by the equation: w^T * x + b = 0, where w is the normal vector to the hyperplane, x is the input feature vector, and b is the bias.

3. **Margin:**

   The margin is the distance between the hyperplane and the nearest data points (support vectors) from each class.

   SVM aims to maximize this margin to improve generalization and reduce overfitting.

4. **Support Vectors:**

   Support vectors are the data points that lie closest to the hyperplane and have the maximum influence on the position and orientation of the hyperplane.

   These points determine the decision boundary and are crucial for defining the margin.

5. **Linearly Separable Data:**

   In the case of linearly separable data, SVM aims to find the hyperplane that separates the classes with the maximum margin.

6. **Non-linearly Separable Data:**

   For non-linearly separable data, SVM uses the kernel trick to map the input features into a higher-dimensional space where the data points become linearly separable.

   Common kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid kernels.

7. **Regularization Parameter (C):**

   SVM introduces a regularization parameter C to control the trade-off between maximizing the margin and minimizing the classification error on the training data.

   A smaller value of C leads to a softer margin, allowing some misclassifications, while a larger value of C leads to a harder margin, penalizing misclassifications more severely.

Logistic Regression:

Logistic Regression is a supervised learning algorithm used for binary classification tasks, where the target variable is categorical and has only two possible outcomes (e.g., 0 or 1, True or False). Here's an overview of the theory behind Logistic Regression:

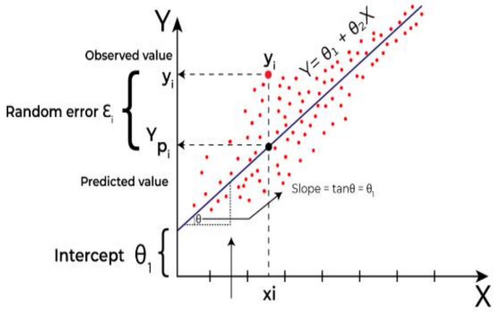


Figure 26

1. Model Representation:

Logistic Regression models the probability that a given input belongs to a particular class.

It uses the logistic function (also known as the sigmoid function) to map input features to probabilities between 0 and 1.

The logistic function is defined as:

2. Hypothesis Function:

The hypothesis function for Logistic Regression is:

This function outputs the estimated probability that

y=1 given input features

x and model parameters

3. Decision Boundary:

Logistic Regression uses a decision boundary to classify inputs into different classes based on their predicted probabilities.

4. Cost Function (Log Loss):

The cost function for Logistic Regression, also known as log loss or cross-entropy loss, is used to measure the error between predicted probabilities and actual class labels.

The cost function is defined as:

The goal is to minimize this cost function to find the optimal parameters.

θ that best fits the training data.

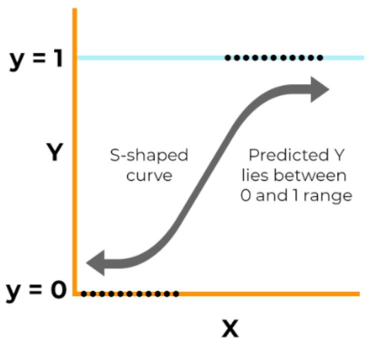


Figure 27

5. Optimization:

Gradient Descent or other optimization algorithms are used to minimize the cost function and find the optimal parameters.

During each iteration of gradient descent, the parameters are updated using the gradient of the cost function with respect to the parameters.

6. Regularization:

Regularization techniques such as L1 regularization (Lasso) or L2 regularization (Ridge) can be applied to Logistic Regression to prevent overfitting and improve generalization performance.

Regularization adds a penalty term to the cost function, which discourages large parameter values.

7. Training and Prediction:

To train a Logistic Regression model, input features and corresponding class labels are used to estimate the model parameters

Once trained, the model can make predictions on new data by computing the probability that each input belongs to the positive class (e.g., class 1) using the learned parameters.

8. Evaluation:

Logistic Regression models can be evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC curve.

These metrics help assess the performance of the model in terms of its ability to correctly classify instances into different classes.

In summary, Logistic Regression is a simple yet powerful algorithm for binary classification tasks, widely used in various fields such as healthcare, finance, and marketing. It models the probability of the positive class using a logistic function and learns optimal parameters through optimization techniques such as gradient descent. Regularization can be applied to prevent overfitting, and the model's performance can be evaluated using various evaluation metrics.

In our research, we have used Linear Regression and Support Vector Regression for **predicting the rating of the two apps using relevant features.**

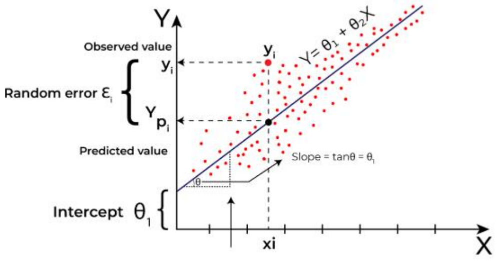


Figure 28

Linear Regression:

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Why Linear regression is Important?

Linear-regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate predictions. Linear regression can be applied to various areas in business and academic study.

Linear regression is used in everything from biological, behavioural, environmental and social sciences to business. Linear-regression models have become a proven way to scientifically and reliably predict the future. Because linear regression is a long-established statistical procedure, the properties of linear-regression models are well understood and can be trained very quickly.

There are two types of Linear Regression:

1. Single Simple Linear Regression

This is the simplest form of linear regression, and it involves only one independent variable and one dependent variable. The equation for simple linear regression is:

- Y is the dependent variable
- X is the independent variable
- β0 is the intercept
- β1 is the slope

2. Multiple Linear Regression

This involves more than one independent variable and one dependent variable. The equation for multiple linear regression is:

- Y is the dependent variable

- X1, X2, …, Xp are the independent variables
- β0 is the intercept
- β1, β2, …, βn are the slopes

The goal of the algorithm is to find the best Fit Line equation that can predict the values based on the independent variables.

In regression set of records are present with X and Y values and these values are used to learn a function so if you want to predict Y from an unknown X this learned function can be used. In regression we have to find the value of Y, So, a function is required that predicts continuous Y in the case of regression given X as independent features.

Best Fit Line

Our primary objective while using linear regression is to locate the best-fit line, which implies that the error between the predicted and actual values should be kept to a minimum. There will be the least error in the best-fit line.

The best Fit Line equation provides a straight line that represents the relationship between the dependent and independent variables. The slope of the line indicates how much the dependent variable changes for a unit change in the independent variable(s).

Support Vector Regression

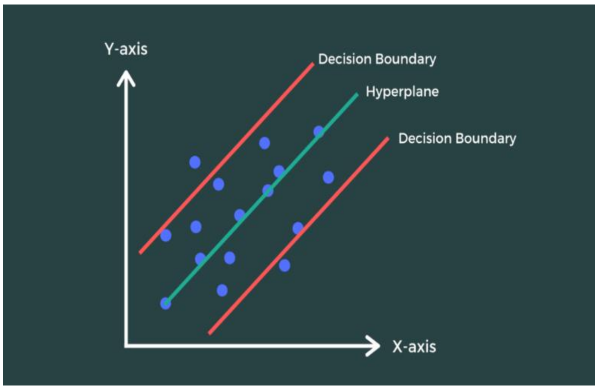


Figure 29

Support Vector Regression (SVR) is a type of machine learning algorithm used for regression analysis. The goal of SVR is to find a function that approximates the relationship between the input variables and a continuous target variable, while minimizing the prediction error.

Unlike Support Vector Machines (SVMs) used for classification tasks, SVR seeks to find a hyperplane that best fits the data points in a continuous space. This is achieved by mapping the input variables to a high-dimensional feature space and finding the hyperplane that maximizes the margin (distance) between the hyperplane and the closest data points, while also minimizing the prediction error.

SVR can handle non-linear relationships between the input variables and the target variable by using a kernel function to map the data to a higher-dimensional space. This makes it a powerful tool for regression tasks where there may be complex relationships between the input variables and the target variable.

Support Vector Regression (SVR) uses the same principle as SVM, but for regression problems.

There are a few important parameters of SVM as explained below:

Kernel: A kernel helps us find a hyperplane in the higher dimensional space without increasing the computational cost. Usually, the computational cost will increase if the dimension of the data increases. This increase in dimension is required when we are unable to find a separating hyperplane in a given dimension and are required to move in a higher dimension:

Hyperplane: This is basically a separating line between two data classes in SVM. But in Support Vector Regression, this is the line that will be used to predict the continuous output

Decision Boundary: A decision boundary can be thought of as a demarcation line (for simplification) on one side of which lie positive examples and on the other side lie the negative examples. On this very line, the examples may be classified as either positive or negative. This same concept of SVM will be applied in Support Vector Regression as well

Support Vector Regression (SVR) extends the principles of Support Vector Machines (SVM) to regression problems, offering a powerful tool for predicting continuous outputs. By leveraging various kernels such as quadratic, radial basis function, and sigmoid, SVR can handle complex and non-linear relationships in the data.

ACCURACY:

As we have mentioned above, for Rating Prediction two models are used Linear regression and support vector regression. To check which model performs better we have used two evaluation metrics Mean squared Error and Mean Absolute Error.

Mean Squared Error (MSE)

Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and predicted values. When a model has no error, the MSE equals zero. As model error increases, its value increases. The mean squared error is also known as the mean squared deviation (MSD).

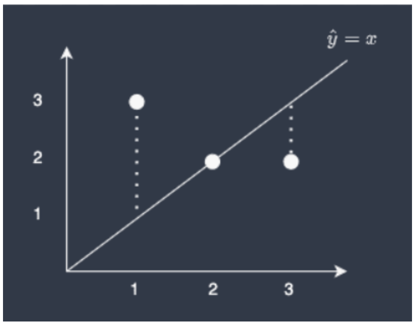


Figure 30

For example, in regression, the mean squared error represents the average squared residual.

Image depicting the relationship between the residuals and the mean squared error.

As the data points fall closer to the regression line, the model has less error, decreasing the MSE. A model with less error produces more precise predictions.

MSE Formula:

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n}$$

Here,

- y_i is the i^{th} observed value.
- $\hat{y}_i$ is the corresponding predicted value.
- n = the number of observations.

<u>Mean Absolute Error:</u>

Absolute Error is the amount of error in your measurements. It is the difference between the measured value and “true” value. This can be caused by your scale not measuring the exact amount you are trying to measure.

Formula:

The formula for the absolute error (Δx) is:

$(\Delta x) = x_i - x$,

Where:

xi is the measurement,

x is the true value.

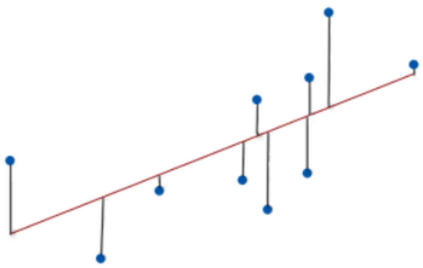


Figure 31

$(\Delta x) = |xi - x|$, The absolute value symbol is needed because sometimes the measurement will be smaller, giving a negative number. On its own, a negative value is fine (-6 just means "six units below") but the problem comes when you're trying to add several values, some of which are positive and some are negative.

Absolute Accuracy Error

Absolute error is also called Absolute Accuracy Error. You might see the formula written this way:

E = Xexperimental – Xtrue.

The Mean Absolute Error (MAE) is the average of all absolute errors.

The formula is:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - x|$$

Where:

n = the number of errors,

Σ = Summation Symbol (which means "add them all up"),

$|x_i - x|$ = the absolute errors.

Precision

Precision is a metric that measures the accuracy of the positive predictions made by the model. It is specifically concerned with the proportion of true positive predictions (correctly identified positive instances) out of all positive predictions made by the model (both true positives and false positives).

Mathematically, precision for a classification model is defined as:

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Here's a breakdown of the terms involved:

True Positives (TP): Instances that are correctly predicted as positive.

False Positives (FP): Instances that are incorrectly predicted as positive (i.e., the model predicts positive, but the actual class is negative).

Precision is particularly important in situations where the cost of false positives is high. For instance, in medical diagnostics, a high precision means that when the model predicts a condition, it is highly likely to be correct, which is crucial to avoid unnecessary treatments or interventions.

AUC-ROC Curve

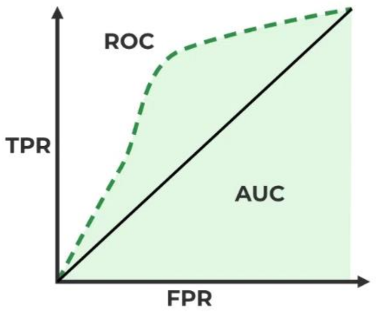


Figure 32

The AUC-ROC curve, or Area Under the Receiver Operating Characteristic curve, is a graphical representation of the performance of a binary classification model at various classification thresholds. It is commonly used in machine learning to assess the ability of a model to distinguish between two classes, typically the positive class (e.g., presence of a disease) and the negative class (e.g., absence of a disease).

8. RESULTS AND INFERENCES

Insightful Perspectives: Exploring Data with PowerBI Data Interpretation

Data Interpretation:

Data interpretation involves analysing and making sense of data to extract meaningful insights and draw conclusions. We have used PowerBI and concluded different Inferences using parameters as **Bivariate and Multivariate Variables** clubbed together in the form of clustered bar charts, maps, graphics etc.

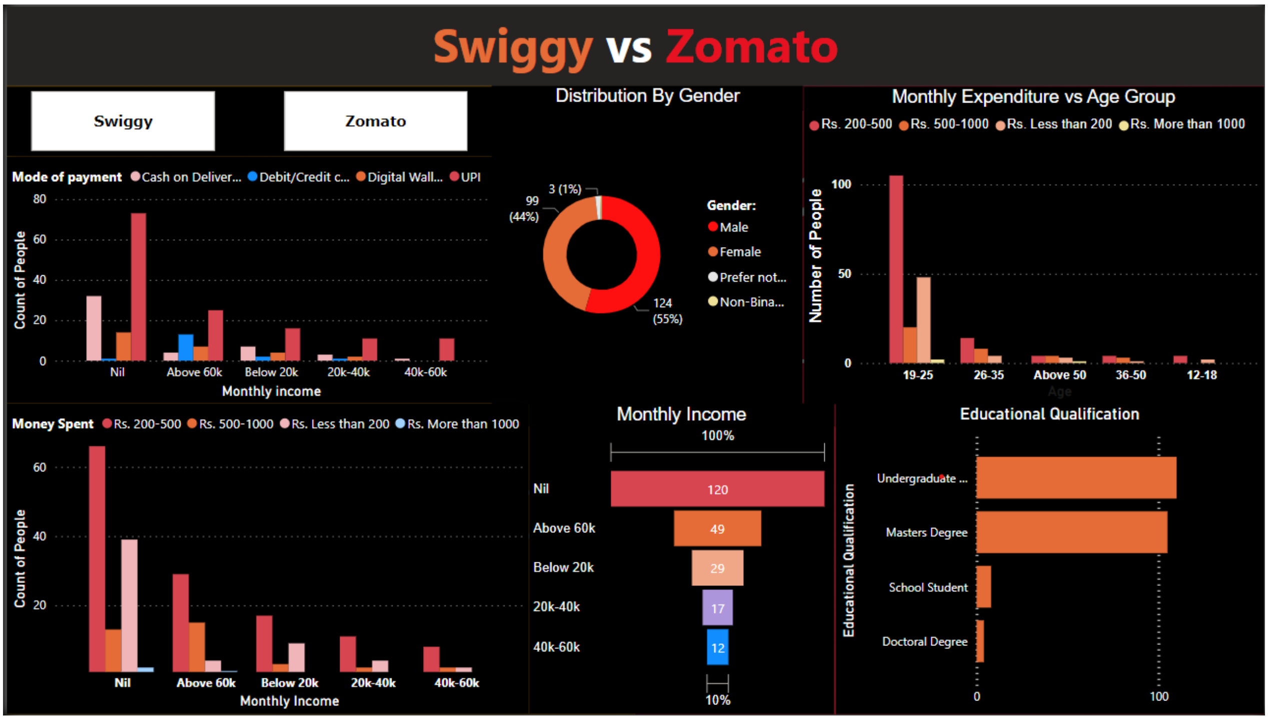


Figure 1 shows,

This is a **Slicer** which **Filters** the data in our report according to our choice of food app.

Figure 2 shows,

This is a **Clustered Column Chart** which is a type of visual representation in Power BI that displays data using vertical columns. It is particularly useful for comparing values across different categories or groups. Here we have shown the **count of people with the comparison of their monthly income with respect to the mode of payment used by them.**

Income -NIL

Mode of Payment	Number of People
UPI	73
COD	32
Digital Wallet	14
Debit/Credit Card	1

Income- Below 20k

Mode of Payment	Number of People
UPI	16
COD	7
Digital Wallet	4
Debit/Credit Card	2

Income - 20k to 40k

Mode of Payment	Number of People
UPI	11
COD	3
Digital Wallet	2
Debit/Credit Card	1

Income- 40k to 60k

Mode of Payment	Number of People
UPI	11
COD	1
Digital Wallet	0
Debit/Credit Card	0

Income -Above 60k

Mode of Payment	Number of People
UPI	25
COD	4
Digital Wallet	7
Debit/Credit Card	13

Figure 3 shows,

This is a **Clustered Column Chart** which is a type of visual representation in Power BI that displays data using vertical columns. It is particularly useful for comparing values across different categories or groups. Here we have shown the **count of people with the comparison of their monthly income with respect to the average amount spent by them.**

Income -NIL

Average Money Spent	Number of People
Less than 200	39
200-500	66
500-1000	13
More than 1000	2

Income -Below 20k

Average Money Spent	Number of People
Less than 200	9
200-500	17
500-1000	3
More than 1000	0

Income - 20k to 40k

Average Money Spent	Number of People
Less than 200	4
200-500	11
500-1000	2
More than 1000	0

Income - 40k to 60k

Average Money Spent	Number of People
Less than 200	2
200-500	8
500-1000	2
More than 1000	0

Income - Above 60k

Average Money Spent	Number of People
Less than 200	4
200-500	29
500-1000	15
More than 1000	1

Figure 4 shows,

This is a **Donut Chart** which is a circular statistical graphic that displays data in the form of a ring across different categories or groups. Here we have shown **different categories of gender who use Swiggy** which includes **Female (47.37%)**, **Male (51.58%)**, **prefer not to say (1%)**.

Gender	Number of People
Male	124
Female	99
Prefer not to say	3

Figure 5 shows,

This is a Funnel chart which shows a sequence of different categories of monthly income.

Monthly Income	Number of People
Nil	120
Below 20k	49
20k - 40k	17
40k - 60k	12
Above 60k	49

Figure 6 shows,

This is a **Clustered Column Chart** which is a type of visual representation in Power BI that displays data using vertical columns. It is particularly useful for comparing values across different categories or groups. Here we have shown the **count of people with the comparison of monthly expenditure of different age groups**

Money Spent per order - Below Rs 200

Age Group	Number of People
12-18	2
19-25	48
26-35	4
36-50	1
Above 50	3

Money Spent per order - Rs 200 to Rs 500

Age Group	Number of People
12-18	4
19-25	105
26-35	14
36-50	4
Above 50	4

Money Spent per order - Rs 500 to Rs 1000

Age Group	Number of People
12-18	0
19-25	20
26-35	8
36-50	3
Above 50	4

Money Spent per order - More than Rs 1000

Age Group	Number of People
12-18	0
19-25	2
26-35	0
36-50	0
Above 50	1

Figure 7 shows,

This **Clustered Bar Chart** is a visual representation of data in which **Rectangular Bars** of different colours or shades are grouped together in categories. Each group of bars represents a distinct category, and within each group, individual bars represent subcategories or different data series. Here we have shown the **Count of people under different categories of Educational Qualification with respect to each app**

Educational Qualification	Number of People
School Student	8
Undergraduate Degree	110
Master’s Degree	105
Doctoral Degree	4

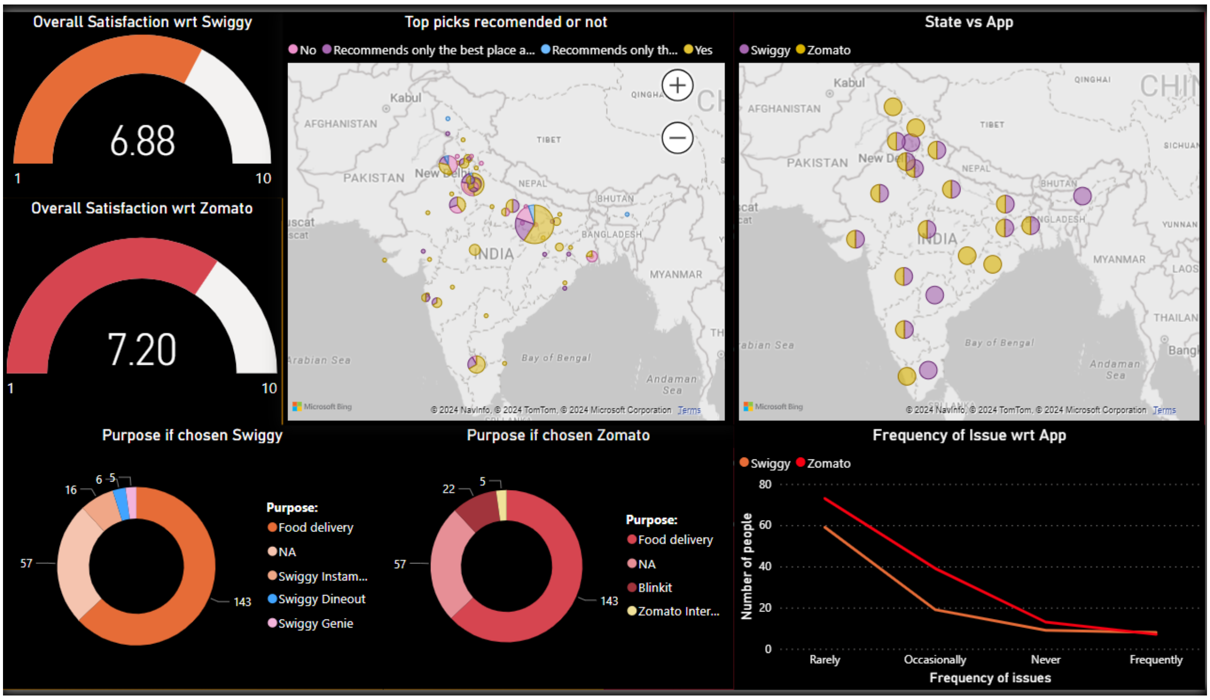


Figure 1 shows,

This is a **Gauge** which **represents a single value within a range**, like a speedometer, to indicate performance against a target. Here it shows the **overall satisfaction of customer with respect to Swiggy where the average value came out to be 6.88**

Figure 2 shows,

This is a **Gauge** which represents a single value within a range, like a speedometer, to indicate performance against a target. Here it shows the **overall satisfaction of customer with respect to Zomato where the average value came out to be 7.20**

Figure 3 shows,

This is a **Donut chart** which is a **circular statistical graphic** that displays data in the form of a ring across different categories or groups. As **Swiggy** is used for **different purposes like**

Food Delivery, Swiggy Instamart, Swiggy Dine Out and Swiggy Genie, we have shown the distribution of the number of people using the above services**.**

Swiggy Services	Number of People
Food Delivery	143
Swiggy Instamart	16
Swiggy Dine Out	6
Swiggy Genie	5

Figure 4 shows,

This is a **Donut Chart** which is a circular statistical graphic that displays data in the form of a ring across different categories or groups. As **Zomato** is used for **different purposes like Food Delivery, Blinkit, Zomato Intercity Legends, we have shown the distribution of the number of people** using the above services.

Zomato Services	Number of People
Food Delivery	157
Blinkit	22
Zomato Intercity Legends	5

Figure 5 shows,

This is a **Line Chart** that displays information as a series of data points connected by straight lines. Here we have shown **number of people having issues with the app may it be Swiggy or Zomato**. We have divided the frequency of issues into four parts shown below:

Frequency of Issues	Zomato	Swiggy
Rarely	73	59
Occasionally	39	19
Never	13	9
Frequently	7	8

Figure 6 shows,

This is a **Map** which typically refers to visualisation that display **Geographical data**. In this map we have shown that if **Swiggy or Zomato** shows the following on their app- ’**Recommends only best place according to the city’, ‘Recommends only famous dishes of the city’, ‘Recommends top picks’, ‘Does not recommend top picks’**. It is shown with the help of the bubble. More the size of the bubble more the count of people.

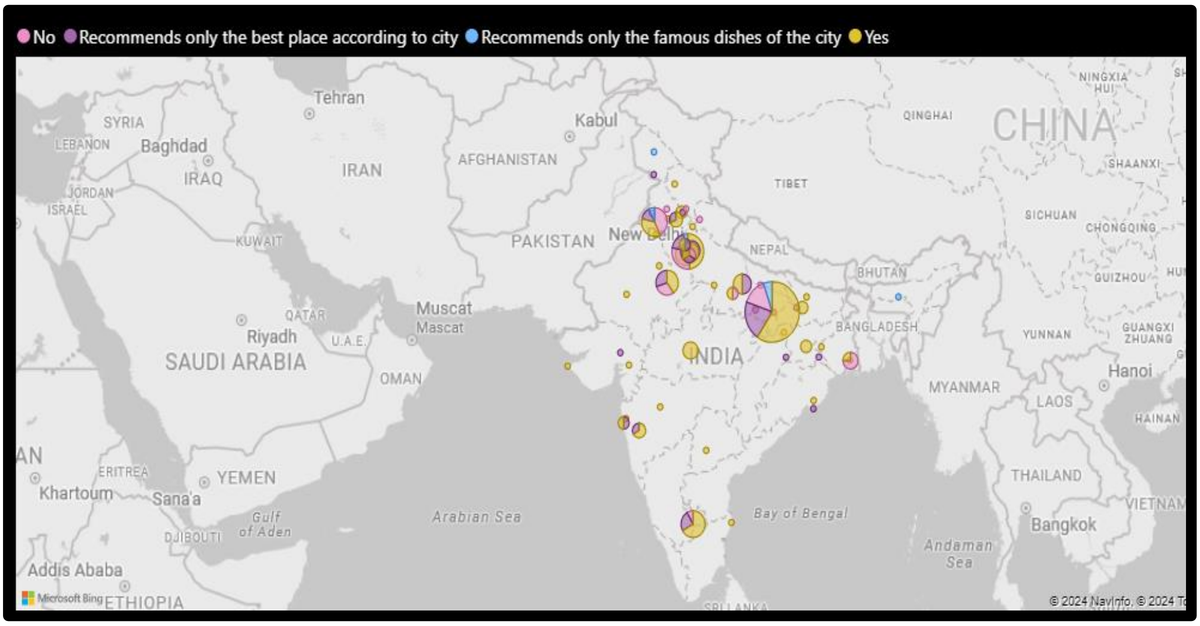


Figure 7 shows,

This is a **Map** which typically refers to visualisation that displays geographical data. In this map we have shown **the dominance of both the apps in different states of India.**

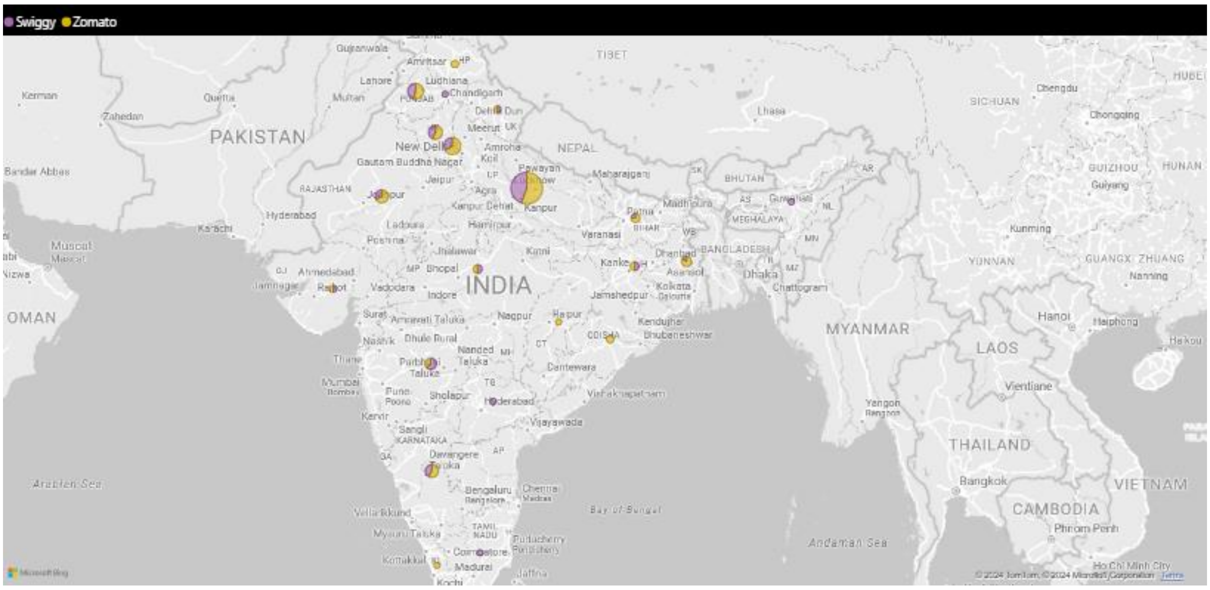


Figure 1 shows,

This is a **Clustered Bar Chart** that displays multiple sets of data using rectangular bars grouped together in clusters. Here we have **shown the dietary preferences of people with respect to the app.**

Satisfaction Level	Zomato	Swiggy
Good	73	59
Fair	39	19
Excellent	13	9
Poor	7	8

Figure 2 shows,

This is a **Stacked Column Chart** that represents multiple data series using vertical bars stacked on top of one another. Here we have shown **the satisfaction level with the condition of food with respect to the app**.

Dietary Preferences	Zomato	Swiggy
Vegetarian	89	56
Non-Vegetarian	38	35
Vegan	0	2
Others	4	2
Gluten Free	1	0

Figure 3 shows,

This is a **Line and Stacked Column Chart** which combines the features of both line and stacked column charts to prevent two different types of data in a single visualisation. Here we have shown **the distribution of the number of people who have issues with the delivery location with respect to the app**.

Frequency of Issues with Delivery Location	Zomato	Swiggy
Yes	69	49
No	63	46

Figure 4 shows,

This is a **Clustered Bar Chart** that displays multiple sets of data using rectangular bars grouped together in clusters. Here we have shown the **usage of different apps counting the number of people having different occupations.**

Occupation	Zomato	Swiggy
Student	87	56
Service	40	33
Business	2	5
Homemaker	3	1

Figure 5 shows,

This is a **Clustered Bar Chart** that displays multiple sets of data using rectangular bars grouped together in clusters. Here we have shown **how many people use Zomato and Swiggy in different cities**.

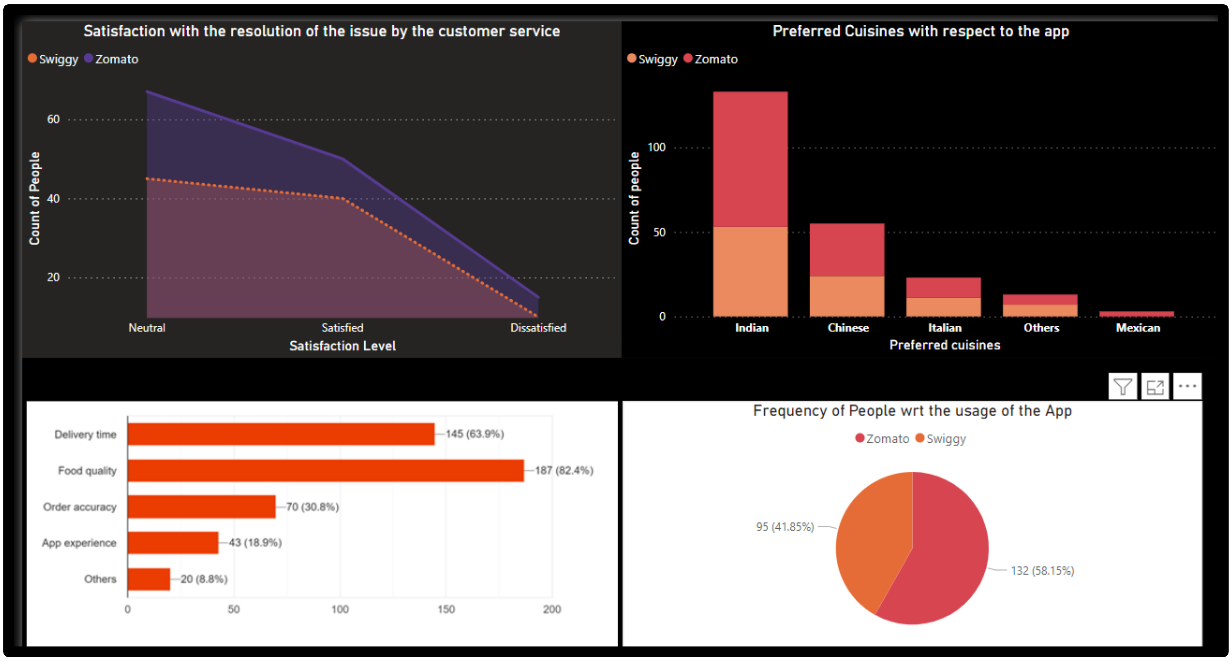


Figure 1 shows,

This is an **Area Chart** that displays data points as coloured areas filling the space between the data points and the horizontal axis. Here we have shown the **satisfaction level with the resolution of the issue by the customer service of Swiggy and Zomato counting the number of people.**

Satisfaction Level	Zomato	Swiggy
Satisfied	50	40
Dissatisfied	15	10
Neutral	67	45

Figure 2 shows,

This is a **Clustered Bar Chart** that displays multiple sets of data using rectangular bars grouped together in clusters. Here we have shown the **factors that customers consider before choosing their app.**

Factors	Number of People
Delivery Time	145
Food Quality	187
Order Accuracy	70
App Experience	43
Others	20

Figure 3 shows,

This is a **Stacked Column Chart** that displays multiple data series as vertical bars stacked on top of one another. This shows the **number of people who have preferred different cuisines from Swiggy and Zomato.**

Preferred Cuisine	Zomato	Swiggy
Indian	80	53
Chinese	31	24
Italian	12	11
Mexican	3	0
Others	6	7

Figure 4 shows,

It is a **Pie Chart** that is a **Circular Statistical Graph** that is divided into **slices** to illustrate numerical proportion. This pie chart **describes how many people use Zomato and how many people use Swiggy.**

Zomato	Swiggy
132	95

Cronbach’s Alpha

The code performs the following steps for each column:

1. Calculates the polarity score for each response in these columns. For example, a response like "Very satisfied" might have a high positive polarity, while "Very dissatisfied" might have a high negative polarity.
2. Computes the mean polarity score for each column, representing the overall sentiment of responses in that column.
3. Stores these polarity scores in a dictionary.

After this, the code calculates Cronbach's alpha using the polarity scores of all columns to measure the internal consistency of the sentiment expressed across these survey questions.

Cronbach's Alpha Output:

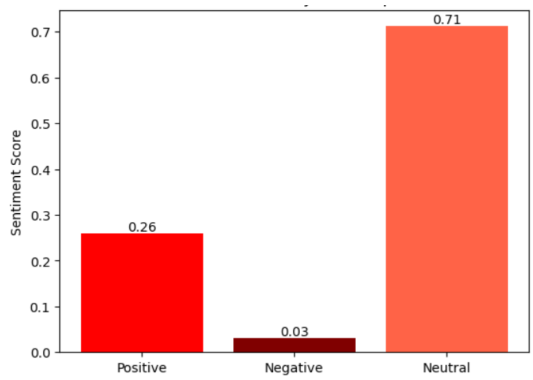
```
Cronbach's Alpha using Polarity Scores: (0.3560387691087516, array([0.23, 0.47]))
```

This value indicates the reliability of the sentiment data. A higher alpha (closer to 1) means the responses are more consistent, while a lower alpha (closer to 0) indicates less consistency.

Sentiment Analysis

Sentiment Scores-Zomato

Sentiment	Sentiment Score
Positive	0.259
Negative	0.03
Neutral	0.712
Compound score	0.994

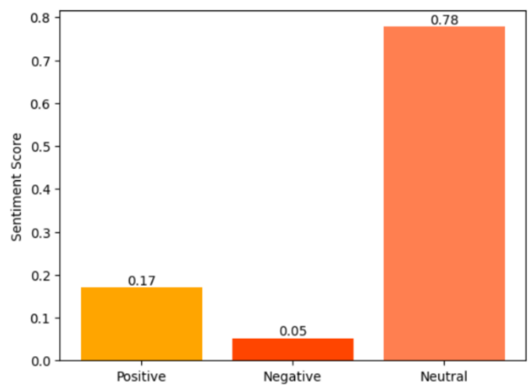


In the above bar chart, we have calculated the sentiment score of open-ended questions asked in the survey form for the online food delivery app Zomato.

Overall Sentiment: Positive

Sentiment Scores-Swiggy

Sentiment	Sentiment Score
Positive	0.17
Negative	0.052
Neutral	0.778
Compound score	0.9998

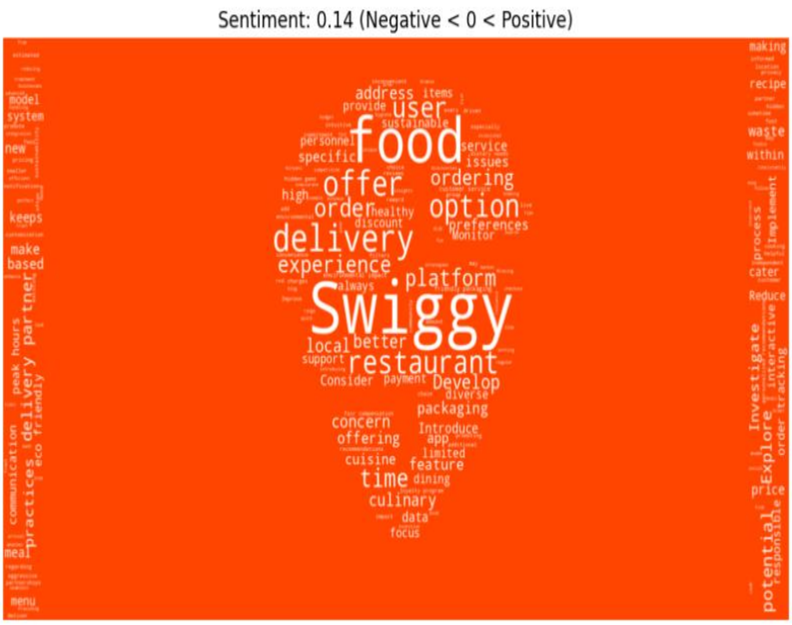


In the above bar chart, we have calculated the sentiment score of open-ended questions asked in the survey form for the online food delivery app Swiggy.

Overall Sentiment: Positive

Word Cloud

In the image the sentiment polarity score of Zomato word cloud is 0.18



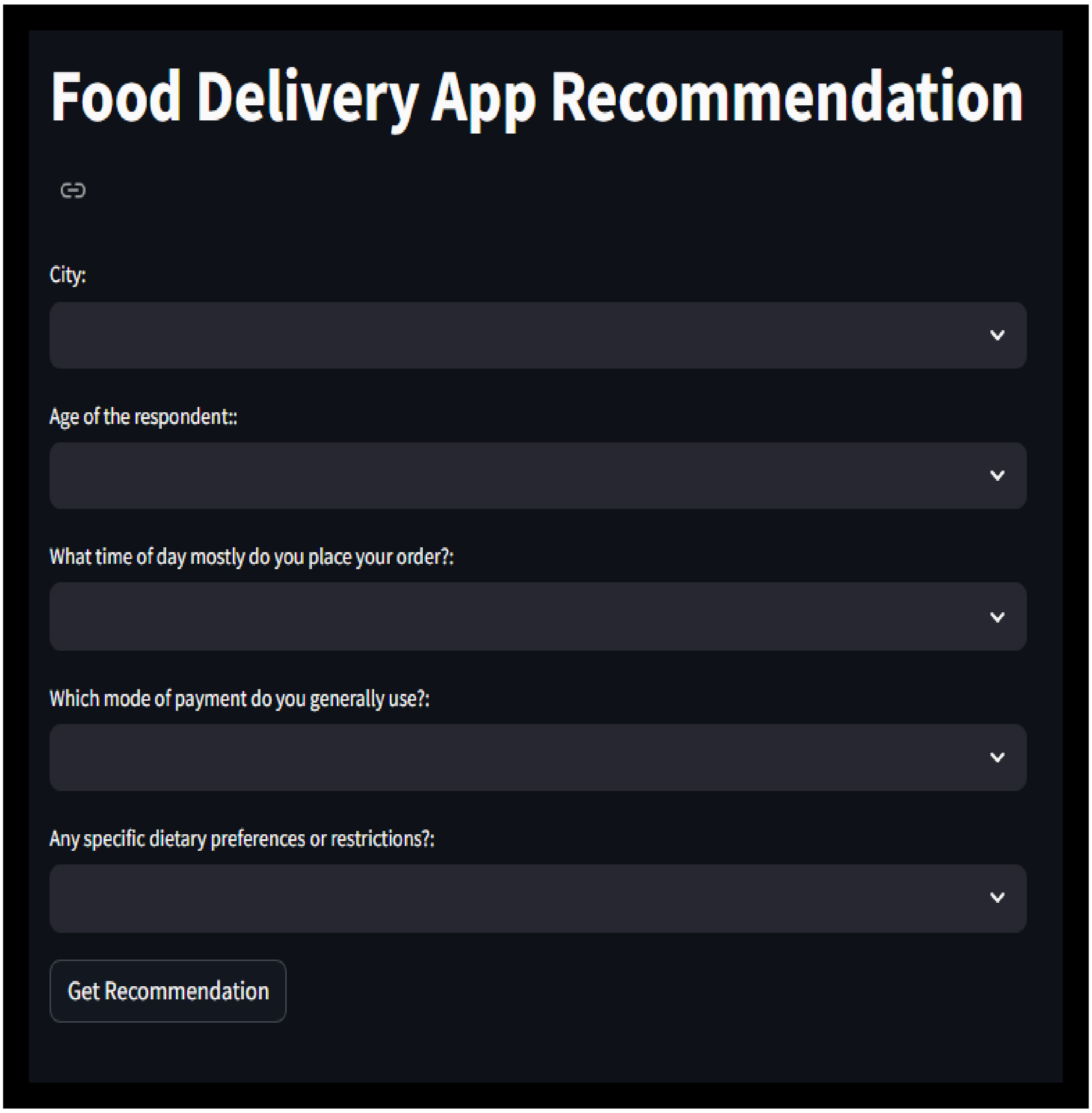
In the image the sentiment polarity score of Swiggy word cloud is 0.14

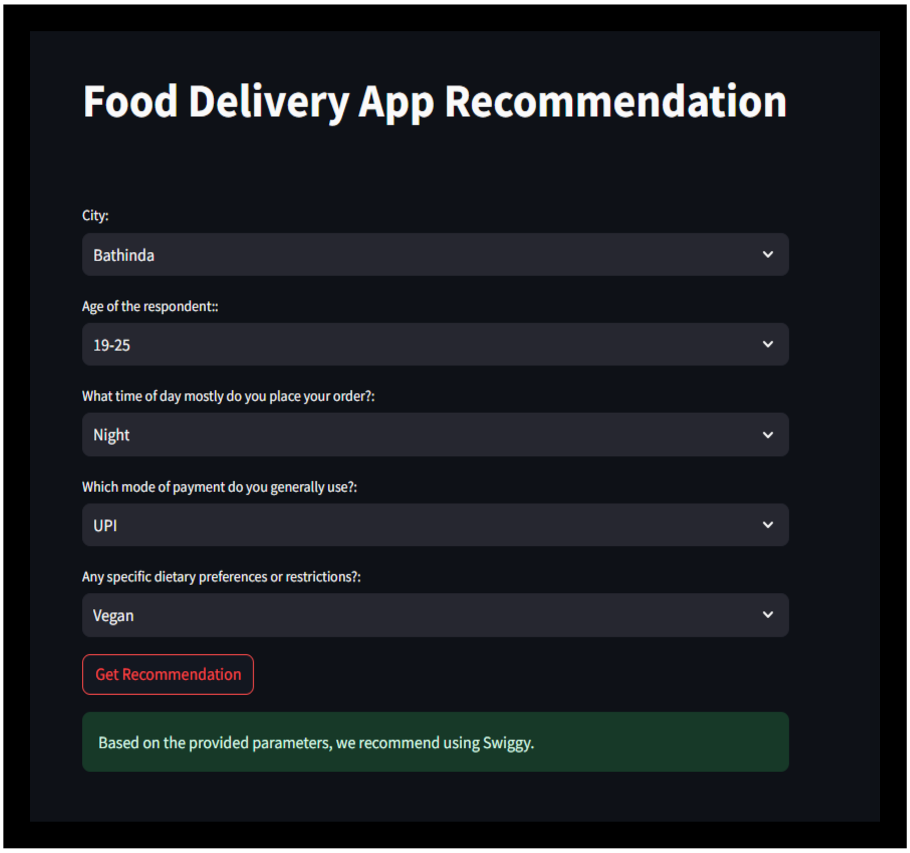
App Recommendation

Our recommendation model takes 5 parameters as input from the new users:

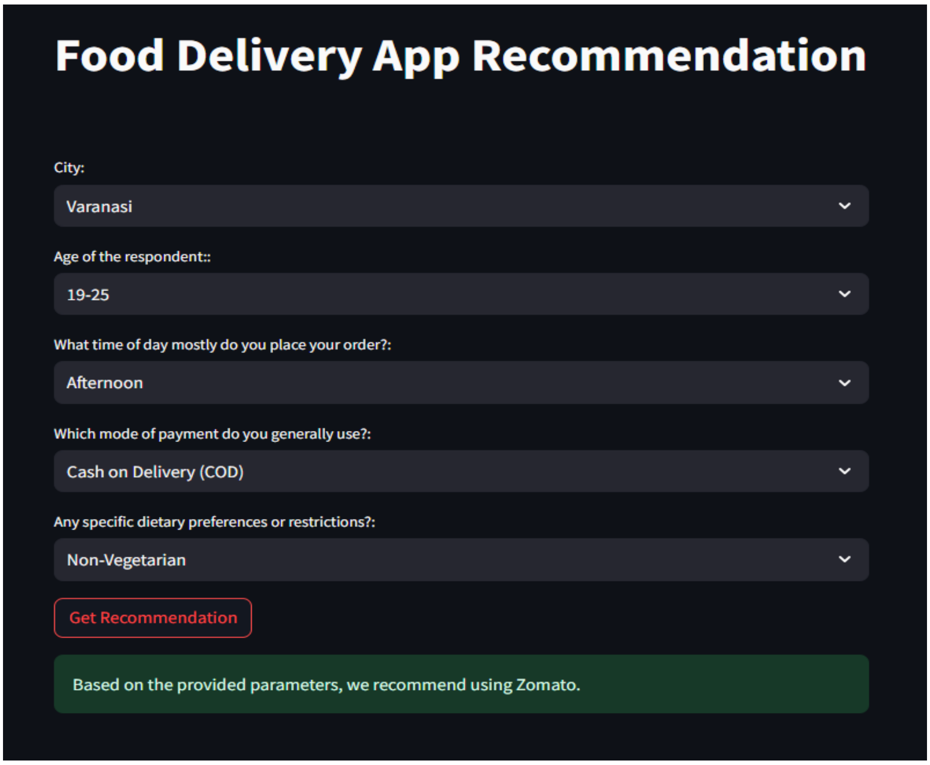
1. City
2. Age of the respondent
3. What time of day mostly do you place your order?
4. Which mode of payment do you generally use?
5. Any specific dietary preferences or restrictions?

We have created a User Interface for above recommendation as shown in Figure below:





In the image, on the basis of parameters the model recommends the app - **Swiggy**



In the image, on the basis of parameters the model recommends the app – **Zomato**

As we can see in the above image the recommendation model pops with 5 questions to be answered by the user. On that basis the model will classify which app to use between Swiggy and Zomato. The model classifies between the two apps on the basis of the data set collected.

Feature Importance for App Prediction

The features of predicting the app are:

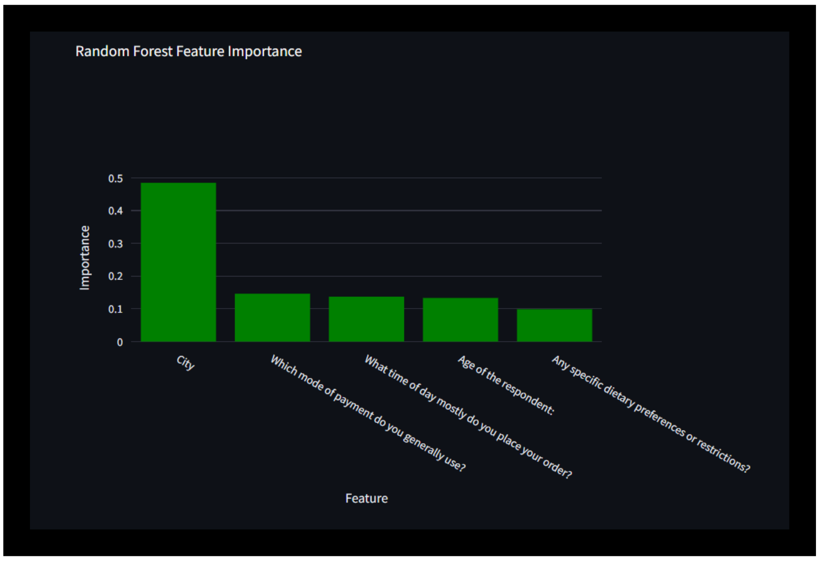
City: May capture location-based preferences and availability of services.

Age of the Respondent: May influence preferences and app usage patterns.

Time of Day: Reflects user habits and demand patterns.

Mode of Payment: Can indicate convenience and preferred transaction methods.

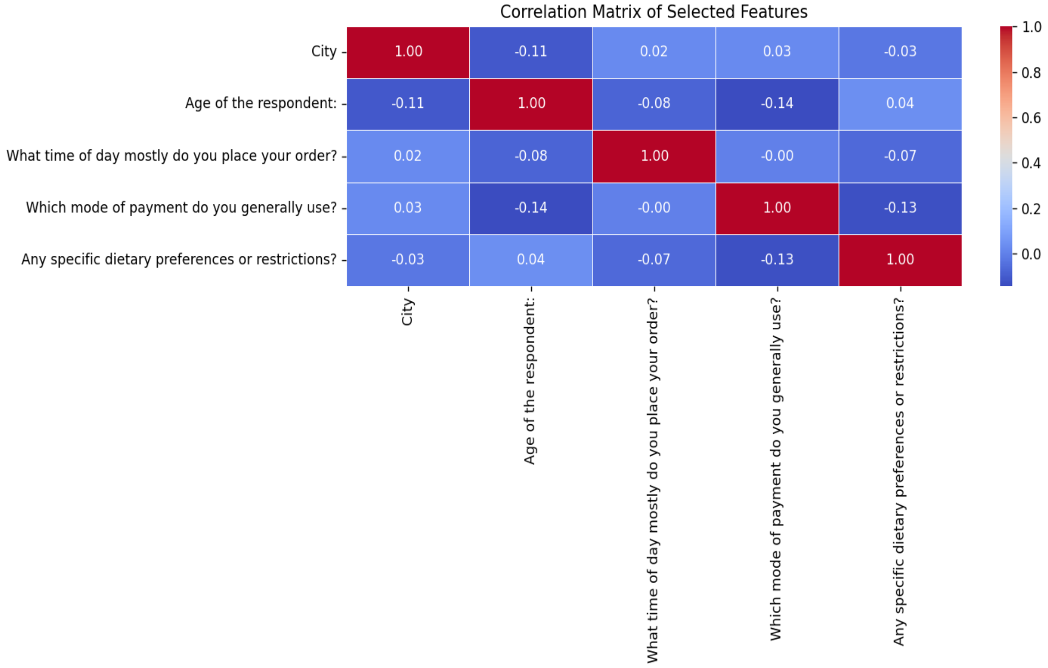
Dietary Preferences or Restrictions: Critical for personalized recommendations and user satisfaction.



Features	Values
Age of the Respondent	0.133
City	0.484
Which mode of payment do you generally use?	0.146
What time of day mostly do you order?	0.137
Any specific dietary preferences or restrictions?	0.098

We can observe that the feature **'City' is 48.4%**, **'Payment Method' is 14.6%**, **'Time of Day' is 13.7%**, **'Age' is 13.3%** and **'Dietary Preferences and Restrictions' is 9.8%** contributing in the prediction of the app.

Correlation Matrix in App Prediction



Feature 1	**Feature 2**	**Correlation Coefficient**
City	Ordering Time	0.02
Mode of Payment	City	0.03
Age	Dietary Preference	0.04

The above table shows the relationship between two features based on correlation coefficient.

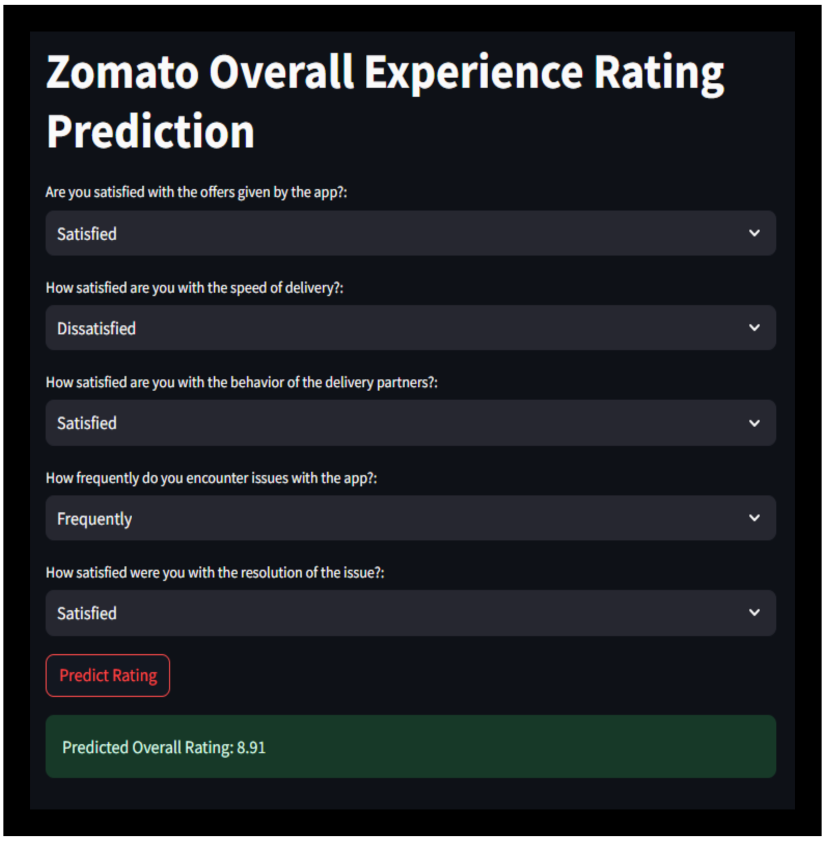
Rating Prediction

Our recommendation model takes 5 parameters as input from the new users:

1. How satisfied are you with the speed of the delivery?
2. Are you satisfied with the offers given by the App?
3. How Satisfied are you with the behaviour of the delivery partner?
4. How frequently do you encounter issues with the App?
5. How Satisfied are you with the resolution of the issue?

We have created a User Interface for above Prediction as shown in Figure below:

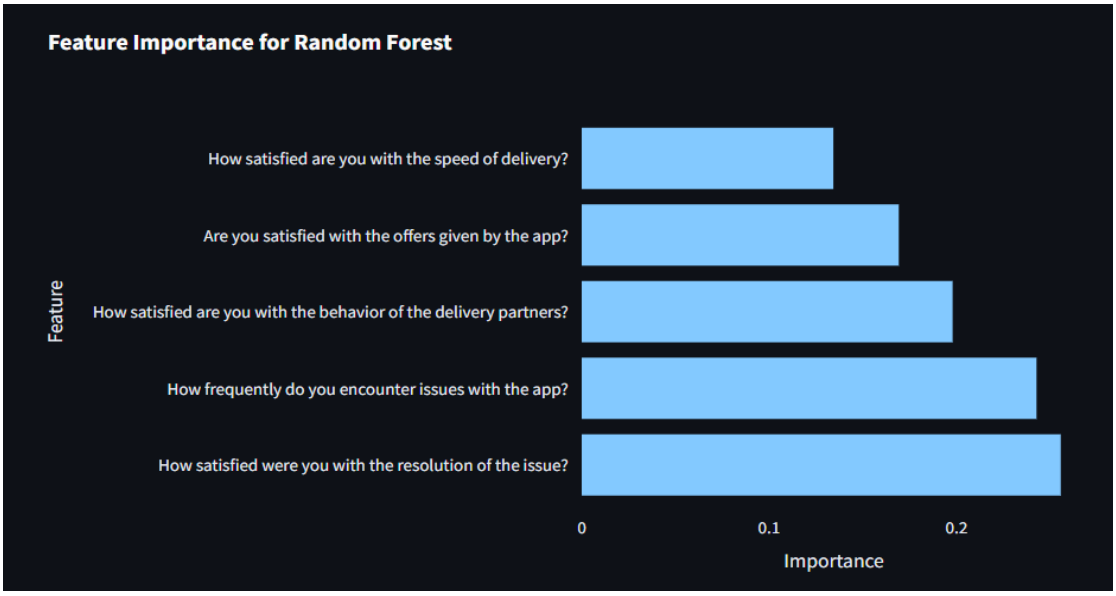
In the above figure, on the basis of parameters the model predicts the Rating of Swiggy App – **7.18**



In the above figure, on the basis of parameters the model predicts the Rating of Zomato App – **8.91**

As we can see in the above image the predictive model pops with 5 questions to be answered by the user. On that basis the model will **predict the overall rating** of the app.

Feature Importance for Rating Prediction



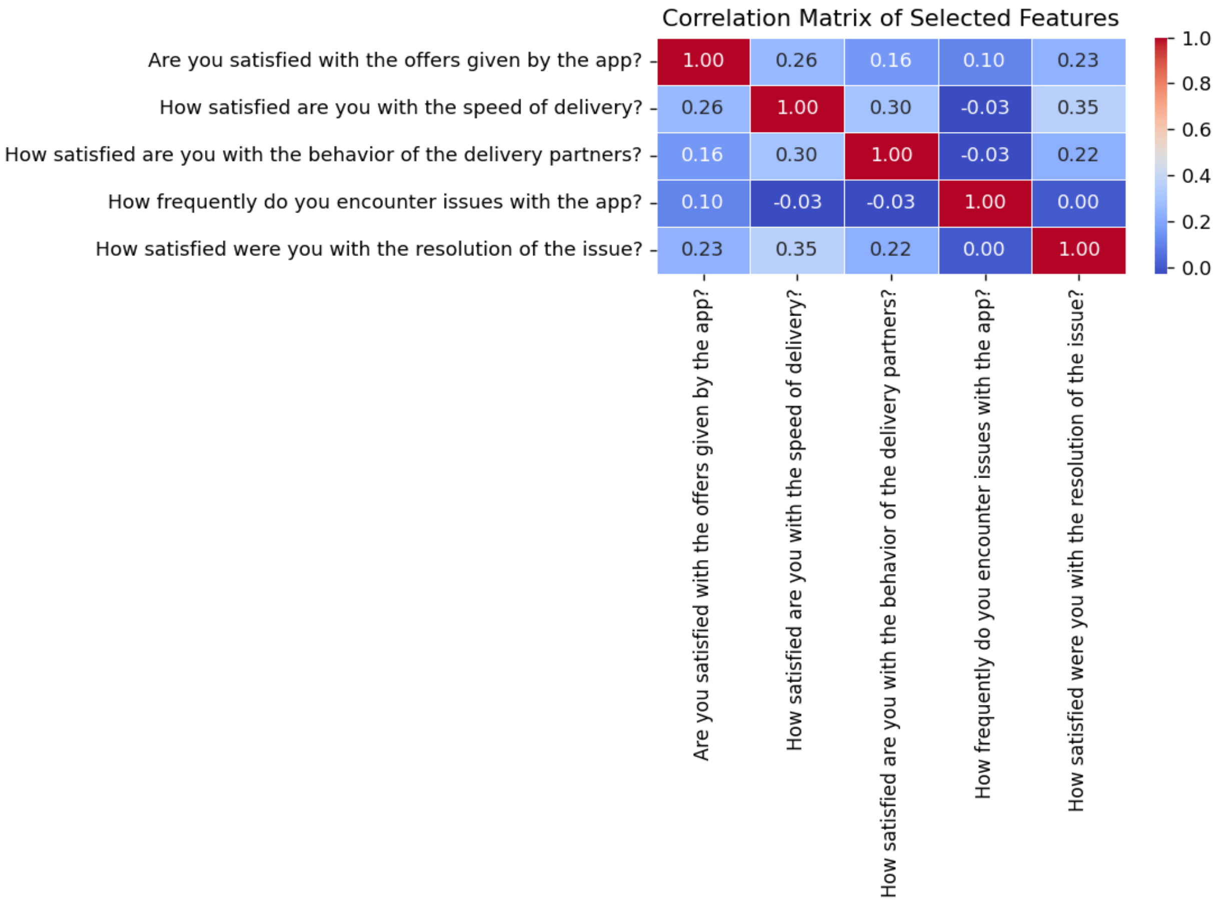
The features for predicting the Rating of the app are:

1. Satisfaction with Offers: Measures how satisfied users are with the app's promotional offers.
2. Satisfaction with Delivery Speed: Measures user satisfaction with delivery timeliness.
3. Satisfaction with Delivery Partner: Behaviour Assesses user satisfaction with the behaviour of delivery personnel.
4. Frequency of App Issues: Measures how often users encounter problems with the app.
5. Satisfaction with Issue Resolution: Evaluates satisfaction with how well issues are resolved by the app's support team.

Features	**Values**
Satisfaction with the resolution of the issue	0.2557
Frequency of encountering issues	0.24
Satisfaction with the behaviour of the Delivery Partner	0.19
Satisfaction with the offers	0.169
Satisfaction with the speed of delivery	0.134

We can observe that the **feature 'Satisfaction with the resolution of the issue' is 25.5%, 'Frequency of encountering issues' is 24%, 'Satisfaction with the behaviour of the Delivery Partner' is 19%, 'Satisfaction with the offers' is 16.9% and 'Satisfaction with the speed of delivery is 13.4%** contributing in the prediction of the app.

Correlation Matrix in Rating Prediction:



Feature 1	**Feature 2**	**Correlation Coefficient**
Satisfied with the delivery speed	Satisfied with the offers	0.26
Satisfied with the behaviour of the delivery partner	Satisfied with the offers	0.16
How frequently do you encounter issues	Satisfied with the offers	0.40
Satisfied with the resolution of the issue	Satisfied with the offers	0.23
Satisfied with the behaviour of the delivery partner	Satisfied with delivery speed	0.30
Satisfied with the resolution of the issue	Satisfied with delivery speed	0.35
Satisfied with the resolution of the issue	Satisfied with the behaviour of delivery partner	0.22

The above table shows the relationship between two features based on correlation coefficient.

However, Correlation is NOT always Causation.

A correlation is a relationship or connection between two variables where whenever one changes, the other is likely to also change. Your growth from a child to an adult is an example. When your height increased, your mass increased, too. And Causation is Getting taller didn’t also make you get wider. Instead, maturing to adulthood caused both variables to increase.

Comparison of Performance of Machine Learning Models used for Recommendation and Rating of the Apps

1. Evaluation Metrics for Rating Prediction

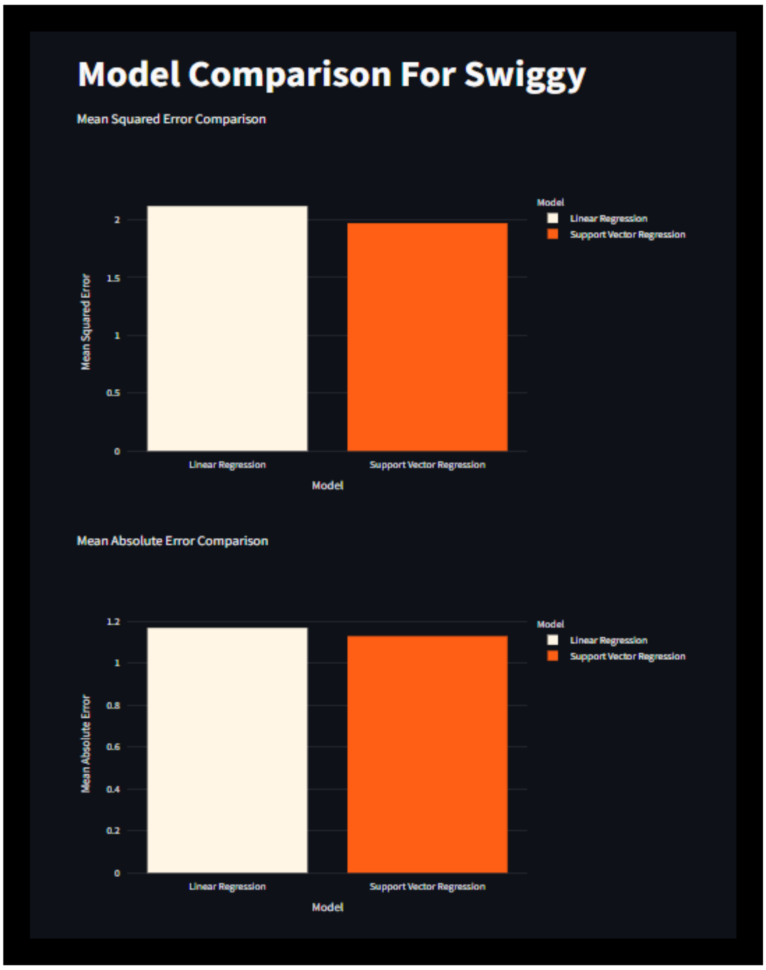
In our research, we have used Linear Regression and Support Vector Regression for **predicting the rating of the two apps using relevant features.**

- **Zomato Accuracy Model Comparison:**

Metric	LR	SVMR
MSE	1.7	1.6
MAE	1.08	1.04

As we know, the error rate predicted by the evaluation metrices used above is **inversely proportional to the performance of the model**, we can conclude that SVMR is a better working machine learning model in comparison to Linear Regression for Rating Prediction. In the above table, **the MSE for linear regression is 1.7 and SVR is 1.6**. Also, **MAE for Linear Regression is 1.08 and for SVR is 1.04**. The height of the bar in the graph does not imply better performance. Hence the above statistics supports the above relationship between error rate and performance.

- <u>**Swiggy Accuracy Model Comparison**</u>



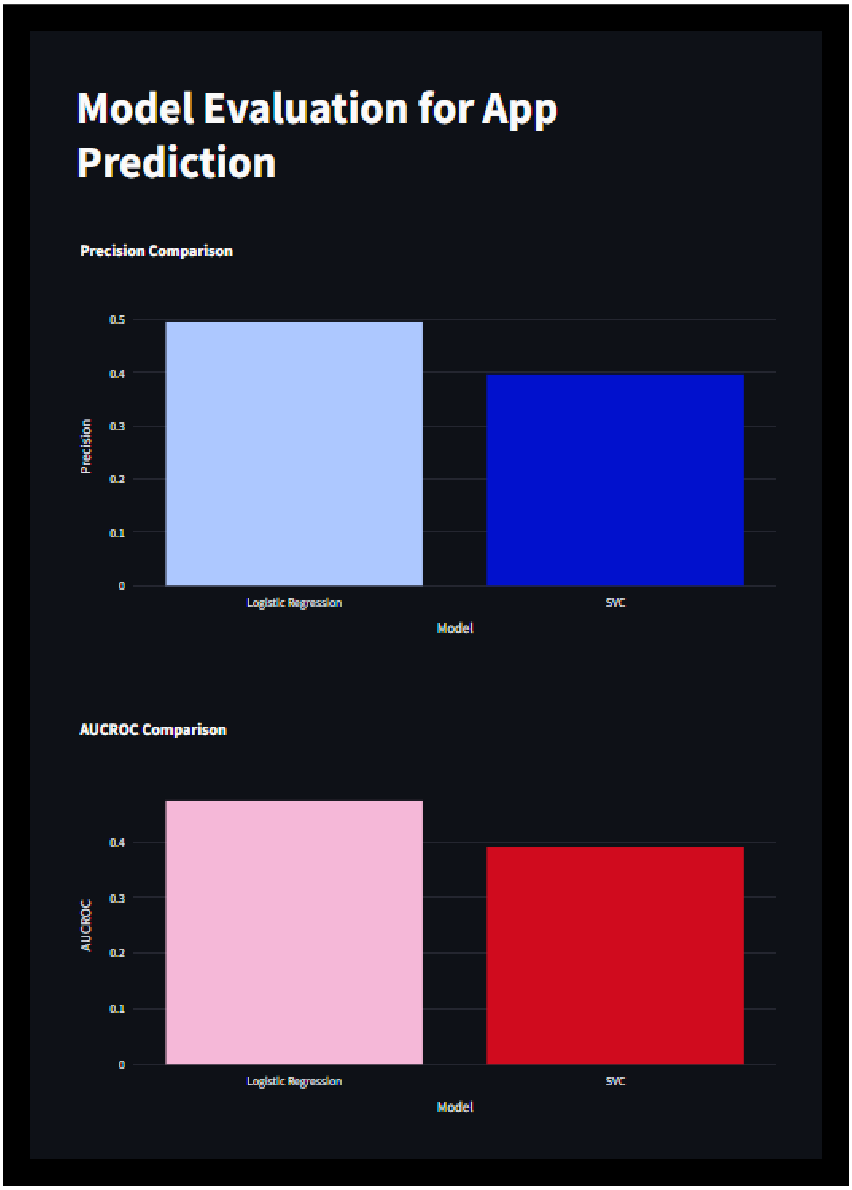
Metric	LR	SVMR
MSE	2.1	1.9
MAE	1.16	1.12

As we know, the error rate predicted by the evaluation metrices used above is **inversely proportional to the performance of the model**, we can conclude that SVMR is a better working machine learning model in comparison to Linear Regression for Rating Prediction. In the above table, **the MSE for linear regression is 2.1 and SVR is 1.9**. Also, **MAE for Linear Regression is 1.16 and for SVR is 1.12**. The height of the bar in the graph does not imply better performance. Hence the above statistics supports the above relationship between error rate and performance.

2. Evaluation Metrics for App Prediction

In our research, we have used Support Vector Machine Classification and Logistic Regression for **classifying between the two apps and then recommending the most suitable app to the user**.

Metric	**LoR**	**SVMC**
Precision	0.496	0.397
AUC-ROC	0.47	0.391



In the above table, **the Precision for Logistic Regression is 0.496 and SVC is 0.397**. Also, **AUC-ROC for Logistic Regression is 0.47 and for SVC is 0. 391.Hence** we can conclude from the above statistics that **Logistic Regression is a better working machine learning model in comparison to Support Vector Machine Classifier** for App Recommendation.

9. CONCLUSIONS

1. According to our research people prefer Zomato (58%) more than Swiggy (42%).
2. The three most important factors that customers consider before choosing their app:
   - Condition of the food after delivery (82.4%)
   - Total time taken for the delivery (63.9%)
   - Accuracy of the order placed (30.8%)
3. The estimated delivery time of Zomato is better than Swiggy. People have said that "Swiggy's estimated time can be inaccurate, leading to long wait times and frustrations."
4. Swiggy's app notification system can be overly aggressive, sending countless pings for irrelevant promotions.
5. Swiggy's surge pricing during peak hours is frustrating and discourages ordering during those times.
6. The newly launched Zomato Augmented reality is a new good feature which is appealing to eyes and is an instrumental way to capture the audience more.
7. Cronbach's Alpha is 0.35 which is between 0 and 1. Higher the value the better is the reliability of our dataset.
8. Using Sentiment Analysis, we have concluded that People have more positive sentiment towards Zomato (25.9%) than Swiggy (17%).
9. For Rating Prediction SVMR is a better working Machine learning model in comparison to linear regression.
10. For App Prediction Logistic regression is a better working model than SVMC.

10.SUGGESTIONS

1. Implement a more reliable table reservation system with fewer technical glitches and booking discrepancies.
2. Develop partnership with farmers and local producers to offer fresh, seasonal ingredients and support sustainable farming practices.
3. Partnership with local transportation services or bike-sharing platforms to offer sustainable and environmentally friendly delivery options.
4. Explore offering advanced features like order customization or group ordering to cater to specific user preferences.
5. The apps should work on minimising the delivery time to some extent keeping in mind the safety of the delivery partner and at the same time providing better customer experience.
6. Both the apps should employ more food delivery partners so that there is no delay in assigning the delivery partners for a specific order.

11.IMPROVEMENTS/FUTURE WORKS

1. Enhance the work already done!
2. Validate our findings by doing the hypothesis testing.
3. Time Series Analysis-Analyse the data and see if there are any changes in the results before covid during covid and after covid.
4. Include more food tech apps/online food delivery apps to make it the comparison between the apps more vivid.

REFERENCES

1. https://www.jetir.org/papers/JETIR2306127.pdf
2. https://www.studocu.com/in/document/pranveer-singh-institute-of-technology/bba/comparison-of-swiggy-and-zomato/63510857
3. https://ijcrt.org/papers/IJCRT2302560.pdf
4. http://dspace.dtu.ac.in:8080/jspui/bitstream/repository/19501/1/Barsha%20Singh_MBA.pdf
5. https://indusedu.org/pdfs/IJREISS/IJREISS_2957_35179.pdf
6. https://medium.com/@ranganathan223/zomato-bangalore-restaurant-analysis-using-power-bi-397fce15f030
7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9355939/
8. https://thebrandhopper.com/2023/05/14/the-swiggy-story-inside-startups-strategy-for-growth-and-expansion/?amp=1
9. https://www.researchgate.net/publication/361393554_Customer_Perception_towards_Online_Food_Delivery_Services-_Development_of_Conceptual_Model
10. https://www.researchgate.net/publication/362761089_A_Study_of_Consumer_behaviour_towards_online_food_delivery
11. https://www.researchgate.net/publication/352477310_LITERATURE_REVIEW_ON_CONSUMER_PERCEPTION_TOWARDS_ONLINE_FOOD_DELIVERY_APPS
12. https://www.researchgate.net/publication/335739176_UNDERSTANDING_CONSUMER_BEHAVIOUR_TOWARDS_UTILIZATION_OF_ONLINE_FOOD_DELIVERY_PLATFORMS
13. https://iaeme.com/MasterAdmin/Journal_uploads/JOM/VOLUME_5_ISSUE_5/JOM_05_05_015.pdf
14. https://www.ijmbs.com/Vol8/issue4/2-suryadev-singh-rathore.pdf
15. https://www.iimspune.edu.in/images/pdf/Journal/Vol2I1-Paper12.pdf
16. https://ijariie.com/AdminUploadPdf/Aditya_Tribhuvan___research_paper_ijariie12316_converted.pdf
17. https://www.ijresm.com/Vol.3_2020/Vol3_Iss3_March20/IJRESM_V3_I3_120.pdf
18. https://ijcrt.org/papers/IJCRT2003375.pdf
19. https://ijcrt.org/papers/IJCRT2104248.pdf

20. https://www.researchgate.net/publication/353622478_REVIEW_ON_CUSTOMER_PERCEPTION_TOWARDS_ONLINE_FOOD_DELIVERY_SERVICES
21. https://ijcrt.org/papers/IJCRT2202443.pdf
22. https://ijrpr.com/uploads/V3ISSUE3/IJRPR3074.pdf
23. https://ijrpr.com/uploads/V3ISSUE5/IJRPR3892.pdf
24. https://www.iimspune.edu.in/images/pdf/Journal/Vol2I1-Paper12.pdf
25. https://ijcrt.org/papers/IJCRT2304637.pdf
26. https://ijcrt.org/papers/IJCRT2304390.pdf
27. https://ijcrt.org/papers/IJCRT2307068.pdf
28. https://onlinelibrary.wiley.com/doi/full/10.1111/ijcs.12877
29. https://www.sciencedirect.com/science/article/pii/S2590291122000134
30. https://www.sciencedirect.com/science/article/abs/pii/S0040162523007783
31. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0290247