Telecom Customer Churn Analysis Using Machine Learning

SUMMER INTERNSHIP REPORT

Submitted by

KARAN PANSURIYA

[220470107131]

In partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

in

Computer Engineering
V.V.P. Engineering College, Rajkot





Gujarat Technological University, Ahmedabad

July -2025

COMPANY CERTIFICATE



25th Jun-2025

TO WHOM IT MAY CONCERN

This is to certify that, Mr./Ms. Karan keshubhai Pansuriya (220470107131) Student of VVP Engineering College (B.Tech-CE), has successfully completed 30 Days Internship in the field of Data Science & Machine Learning during the period of 17th June-2025 to 16th July-2025 During the period of this internship program with us, student of your College/University exposed to different processes and found sincere and hardworking.



With Best Regards,

Mr. Sagar Jasani

CEO.

Brainybeam Info-Tech PVT LTD

Ahmedabad (Gujarat-INDIA)







V.V.P Engineering College

Virda-vajdi Rajkot.

(Affiliated with GTU)

CERTIFICATE

This is to certify that the summer internship report submitted along with the project entitled "Telecom Customer Churn Analysis Using Machine Learning" has been carried out by KARAN PANSURIYA (220470107131) under my guidance, in partial fulfilment of the course Summer Internship (3170001) for the 7th Semester of B.E. in Computer Engineering under Gujarat Technological University, Ahmedabad, during the academic year 2025–26.

Prof. Kinjal DaveInternal Guide

Dr. Tejas PataliaHead of the Department





ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of this task would be incomplete without the mention of the people who made it possible, whose constant guidance, support and encouragement crown all the efforts with the success.

I would like to sincerely thank the Head of the Department, Dr. Tejas Patalia, for his invaluable guidance and support during the duration of this internship. I am also deeply grateful to **Prof. Kinjal Dave** for her consent to be my guide and for his constant encouragement throughout this journey. Last but not least, I would like to thank my industry mentor **Mr. Amit Tiwari**, who helped me a lot during this project. My sincere dedication and keen desire to learn something new helped me achieve success in the project.

With Sincere Regards
KARAN PANSURIYA

ABSTRACT

This project aims to develop a Telecom Customer Churn Analysis System employing advanced data analytics and machine learning techniques. The primary goal is to build an effective model for predicting customer churn in the telecom industry, enabling proactive customer retention strategies and improved service quality.

The dataset utilized in this project comprises diverse customer-related attributes such as usage patterns, account information, service features, and customer demographics. Exploratory data analysis (EDA) is conducted to understand the relationships between these factors and customer churn.

Machine learning algorithms are employed to predict customer churn. Various models including logistic regression, decision trees, random forest, gradient boosting, and neural networks are trained on historical customer data, considering factors that influence customer behavior and retention.

The project aims to provide a user-friendly interface allowing telecom providers to input customer attributes and receive churn predictions. This system aids in customer relationship management, resource allocation, and service enhancements, ultimately reducing customer turnover and improving business sustainability.

List Of Figures

List of Figures	Page Number
Figure 2.3.1 An interactive line plot	7
Figure 2.3.2 A combined visualization comprising both a bar chart and a pie chart	8
Figure 2.3.3 A combined visualization comprising both a box plot and a histogram	8
Figure 2.3.4 Correlation heatmap	10
Figure 2.3.5 Clustering results	11
Figure 2.3.6 Gender churn distribution	11
Figure 2.3.7 Data Processing	12
Figure 2.3.8 Evaluating model	12
Figure 2.4.1 Gain Chart	13

Symbol and Abbreviations

E-R Diagram:

Entity
Decision make
 Line of flow
Relation Ship

Data Flow Diagram:

	User sub process
	Initial activity
\longrightarrow	Control flow

Class Diagram:

Circle	← Class Name	Class
- x-coord - y-coord # radius	← Attributes	
+ findArea() + findCircumference() + scale()	← Operations	
\longrightarrow		Direct Association
•>		Dependency

Use Case Diagram:

	System Boundary
«uses»	Extends of relation
usecase	User action state
user	User
Object1	Represent Object
Message1	Message flow information
message call	Message call it self

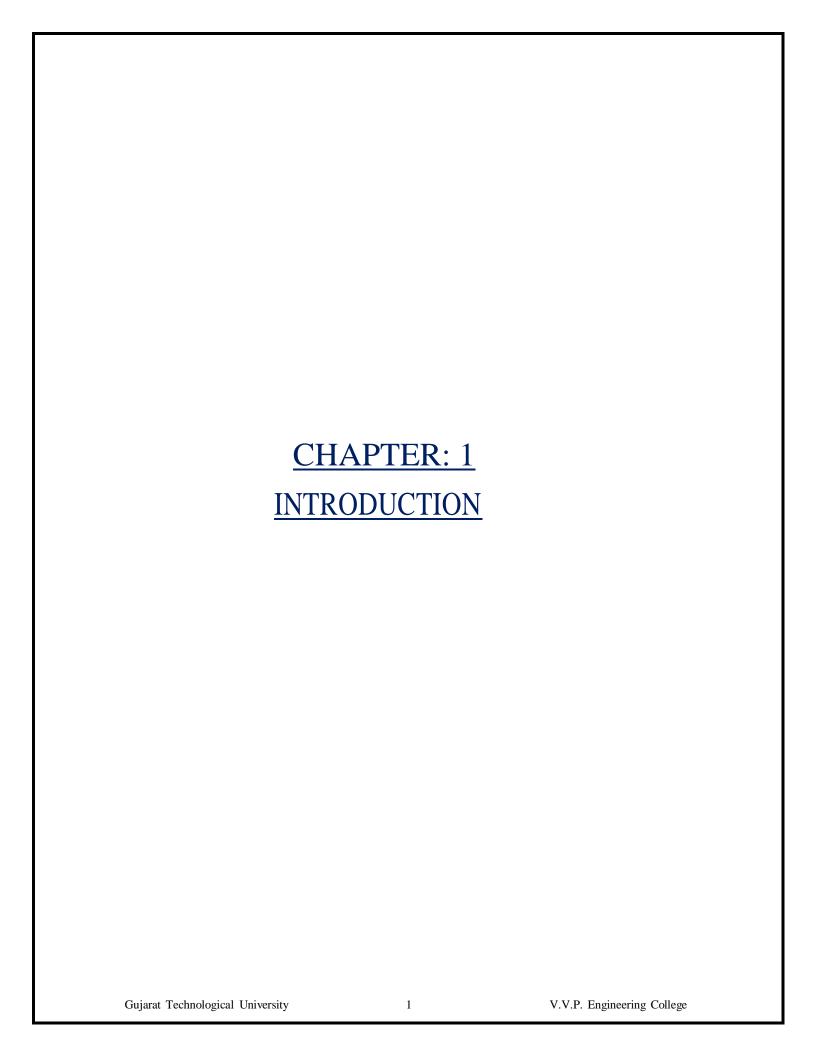
Activity Diagram:

	Initial state
\longrightarrow	Flow of control
State1	Initial activity
ActionState1	Sub activity
\Diamond	Decision making
	End of process

Table of Contents

Title Page	
Completion Certificate.	
College Certificate	i
Acknowledgement	ii
Abstract	iii
List of Figures	iv
Symbol and Abbreviations	vi
Title Page. Completion Certificate i Completion Certificate i Acknowledgement ii Abstract iii List of Figures iv Symbol and Abbreviations vi Table of Contents vii Chapter 1 Introduction 1 1.1 Objectives of Internship 2 1.2 About the company 2 1.3 About Technology 3 1.4 Project Information 3 Chapter 2 Roles and Responsibilities during internship 5 2.1 Problems Given 6 2.2 My Roles and Responsibilities 6 2.3 Daily Tasks and Activities 7 2.4 Internship Schedule (Gantt Chart) 9 Chapter 3 Skills Learned 14 3.1 About the skill/s 15 3.2 How do I learnt the skill/s 16 Chapter 4 Overall Experience 18 4.1 Technical Experience 19	
Chapter 1 Introduction	1
1.1 Objectives of Internship	2
1.2 About the company	2
1.3 About Technology	3
1.4 Project Information	3
Chapter 2 Roles and Responsibilities during internship	5
2.1 Problems Given	6
2.2 My Roles and Responsibilities	6
2.3 Daily Tasks and Activities	7
2.4 Internship Schedule (Gantt Chart)	9
Chapter 3 Skills Learned	14
3.1 About the skill/s	15
3.2 How do I learnt the skill/s	16
Chapter 4 Overall Experience	
4.1 Technical Experience	19
4.2 Personal Experience.	20

Chapter 5 Conclusion	21
5.1 Conclusion	22
Chapter 6 Future Scope	23
6.1 Future Scope	24
Chapter 7 Bibliography	
7.1 Bibliography	26
Plagiarism Report	



1.1 Objectives of Internship:

The primary objective of the internship was to bridge the gap between academic learning and industry practices by providing real-world exposure to software development concepts, tools, and workflows. This internship aimed to equip students with the technical skills and practical experience required to excel in the IT industry.

The specific objectives of the internship were as follows:

- To gain hands-on experience in industry-relevant technologies and tools.
- To understand the fundamentals of backend development and cloud-based solutions.
- To apply theoretical knowledge to practical scenarios through project-based learning.
- To develop critical thinking and problem-solving abilities in real-world contexts.
- To improve communication, collaboration, and professional work ethics in a virtual environment.
- To complete structured learning modules and assessments for skill certification.

1.2 About The Company:

BrainyBeam Info Tech Pvt. Ltd. is a technology company dedicated to empowering businesses
with innovative solutions in data analytics, artificial intelligence, and machine learning. With
a mission to drive actionable insights and enhance decision-making processes, BrainyBeam
controls the power of advanced technologies to solve complex challenges across various
industries.

• Vision:

Our vision at BrainyBeam is to revolutionize the way businesses leverage data to achieve their goals. We strive to be at the forefront of technological advancements, delivering impactful solutions that drive growth, efficiency, and competitive advantage for our clients.

• Mission:

At BrainyBeam Info Tech Pvt. Ltd., our mission is to empower organizations with data-driven intelligence that enables them to unlock new opportunities, optimize processes, and stay ahead in today's dynamic market landscape.

• About Technology:

Front-End: Not Applicable

Back End: Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn

Database

1.3 Project Information:

• Project Title:

Telecom Customer Churn Analysis Using Machine Learning

• Project Overview:

- Focuses on predicting customer churn in the telecom industry.
- Uses **data analytics and machine learning** to identify churn patterns.
- Analyzes customer demographics, usage behavior, and service features.
- Builds predictive models like logistic regression, decision trees, random forest, gradient boosting, and neural networks.
- Provides a **user-friendly interface** for churn prediction.
- Helps telecom companies in customer retention, service improvement, and decisionmaking.

• Impact:

- Enables telecom companies to reduce customer turnover.
- Improves customer relationship management and service quality.
- Optimizes resource allocation for retention strategies.
- Enhances business sustainability through data-driven decision-making.

• Learning Outcomes:

- Practical experience in data preprocessing, feature engineering, and exploratory data analysis (EDA).
- Hands-on knowledge of machine learning models (logistic regression, decision trees, random forest, gradient boosting, neural networks).
- Understanding of **model evaluation metrics** for classification problems.
- Skills in developing a **user-friendly predictive system** integrating ML with real-world applications.
- Exposure to business problem-solving using analytics.

CI ROLES AND RE	HAPTER: 2	_	
	NTENSHIP		
Gujarat Technological University	5	V.V.P. Engineering College	

2.1 Problems Given:

- During the internship/project, the following key problems were identified and given as challenges to solve:
 - Telecom companies face significant customer turnover, leading to revenue loss and reduced market share.
 - Companies struggle to identify customers at risk of leaving before it happens.
 - Multiple factors like demographics, usage patterns, and service quality make it difficult to manually analyze churn risk.
 - Without accurate prediction models, companies spend resources on generalized retention campaigns instead of targeting the right customers.

2.2 My Roles and Responsibilities:

Data Collection & Preprocessing

Gathered telecom datasets and performed data cleaning, handling missing values, and feature engineering.

• Exploratory Data Analysis (EDA)

Analyzed customer demographics, service usage, and account features to identify churn patterns and key influencing factors.

• Model Development

Implemented machine learning algorithms such as logistic regression, decision trees, random forest, and gradient boosting.

Optimized models and compared performance metrics to select the best one.

Reporting & Documentation

Prepared reports summarizing analysis, insights, and recommendations.

Documented methodology, model evaluation, and system workflow.

2.3 Daily Tasks and Activities:

Day	Task / Activity Description
Day 1- 2	A summary of Data Science using Python and Generate a 2D array using NumPy and scale it by multiplying it with a single value.
Day 3-4	 Practiced NumPy basics (array operations) and Worked on reshaping and transposing arrays.
Day 5-6	Loaded and inspected data using Pandas and Removed duplicate rows from datasets.
Day 7-8	- Applied filtering techniques on data and Renamed dataset columns.
Day 9-10	Created line plots for visualization and Designed bar charts for categorical data.
Day 11-12	Implemented box plots for data distribution analysis and Created heatmaps to analyze correlations.
Day 13-14	- Visualized population trends using Seaborn line plots.
Day 15-16	- Conducted scatterplot analysis of correlation.
Day 17-18	 Generated histograms to analyze quantitative variables and Performed box plot analysis of central tendency and dispersion.
Day 19-20	- Worked on exploratory practice exercises and Continued visualization and EDA tasks.
Day 21-22	- Feature extraction using Linear Discriminant Analysis (LDA).
Day 23-24	 Feature extraction from text using TF-IDF and Explored Latent Dirichlet Allocation (LDA) for topic modeling.
Day 25-26	 Rephrasing and feature extraction using TF-IDF and Normalized data using Z-Scores.

Day 27-28	Normalized data using Min-Max scaling and Compared different data normalization techniques.
Day 29-30	- Worked on implementing machine learning models and Practiced model comparison, tuning, and evaluation.

Table 2.3.1 Daily Task and Activities

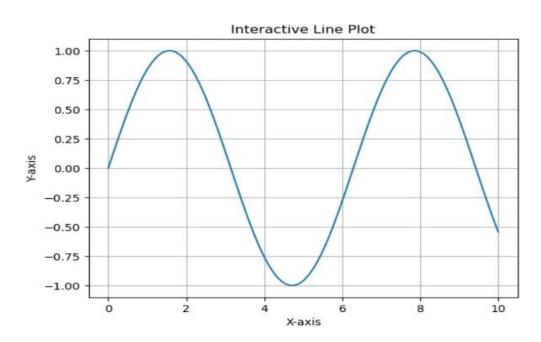


Fig:2.3.1 An interactive line plot showing the sine function.

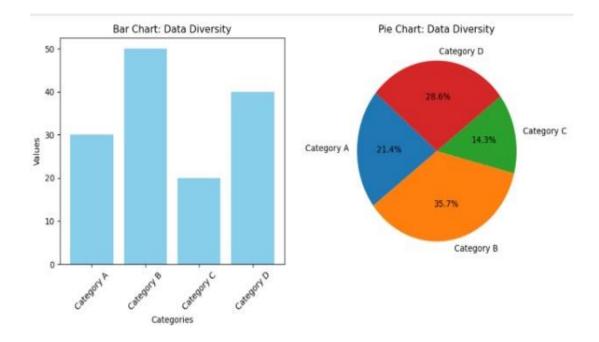


Fig:2.3.2 A combined visualization comprising both a bar chart and a pie chart, providing a comprehensive view of data diversity.

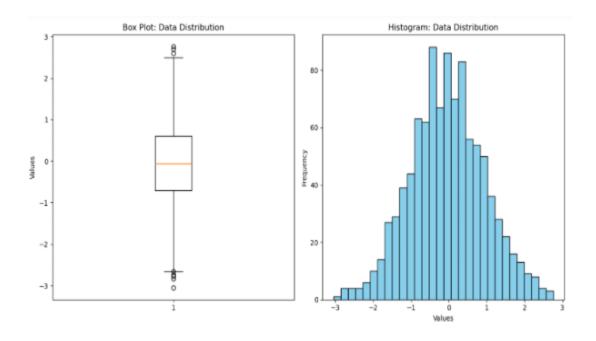


Fig:2.3.3 A combined visualization comprising both a box plot and a histogram, providing a comprehensive analysis of data distribution.

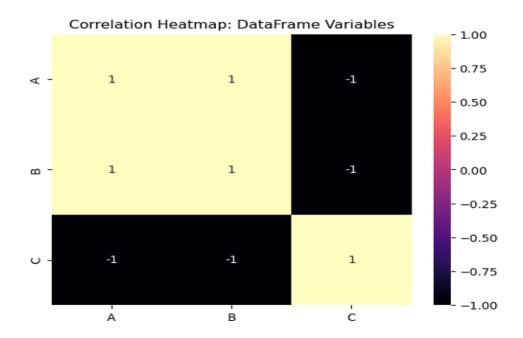
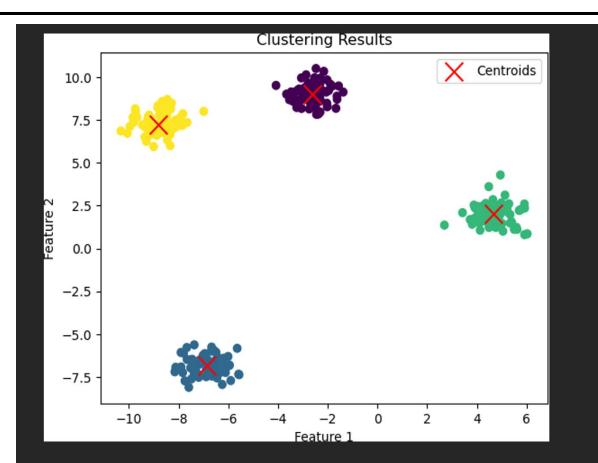


Fig:2.3.4 A correlation heatmap illustrating the relationships between variables in the DataFrame.



The output of the code is a scatter plot displaying the clustering results. Data points are colored according to their assigned clusters, and centroids are marked with red crosses. This visualization helps in identifying hidden structures in the dataset and understanding the grouping patterns discovered by the clustering algorithm.

Fig:2.3.5



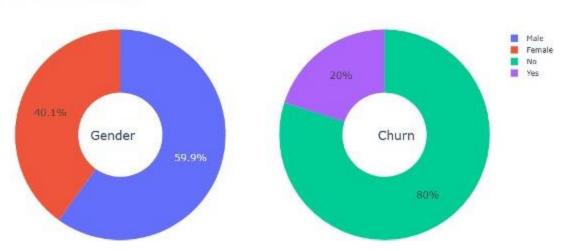


Fig:2.3.6

one-hot encoding

```
In [25]: columns_to_encode = ['telecom_partner', 'gender', 'state', 'city']
         df_encoded = pd.get_dummies(df, columns=columns_to_encode)
         print(df_encoded.head())
            customer_id age pincode date_of_registration num_dependents
                          25 755597
55 125926
57 4220
                                         2020-01-01
2020-01-01
                     1 25
2 55
3 57
                                                  2020-01-01
                                                                             0
                                         2020-01-01
2020-01-01
                           46
                                522841
            estimated_salary calls_made sms_sent data_used churn ... \
                                  44 45 -361
62 39 5973
49 24 193
                       124962
130556
                                                                       9 ...
                       148828
                        38722
                                        80
                                                           9377
                       55098
            state_Tripura state_Uttar Pradesh state_Uttarakhand state_West Bengal \
                                False
         0
                     False
False
                                                               False
False
                                                                                   False
                     False
                                           False
                                                               False
                     False
                                           False
                                                               False
                                                                                   False
            city_Bangalore city_Chennai city_Delhi city_Hyderabad city_Kolkata \
                                    False False
False True
False False
False True
False True
         0
                      False
                                                            False
False
                                                                                 True
False
                      False
                      False
                                                                   False
                                                                 raise
False
                      False
                                                                                  True
                      False
                                    False
                                                                                 False
            city_Mumbai
         0
                   False
                    True
                   False
                   False
                   False
```

Fig:2.3.7 Data Preprocessing

```
Training Mean Squared Error: 0.16
```

Training RMSE: 0.4 Training MAE: 0.32

Validation Mean Squared Error: 0.16

Validation RMSE: 0.4 Validation MAE: 0.32

Test Mean Squared Error: 0.16

Test RMSE: 0.4 Test MAE: 0.32

Fig:2.3.8 Evaluating Model

2.4 Internship Schedule (Gantt Chart)

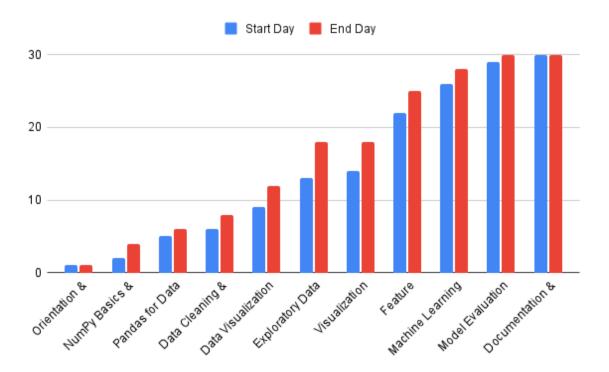
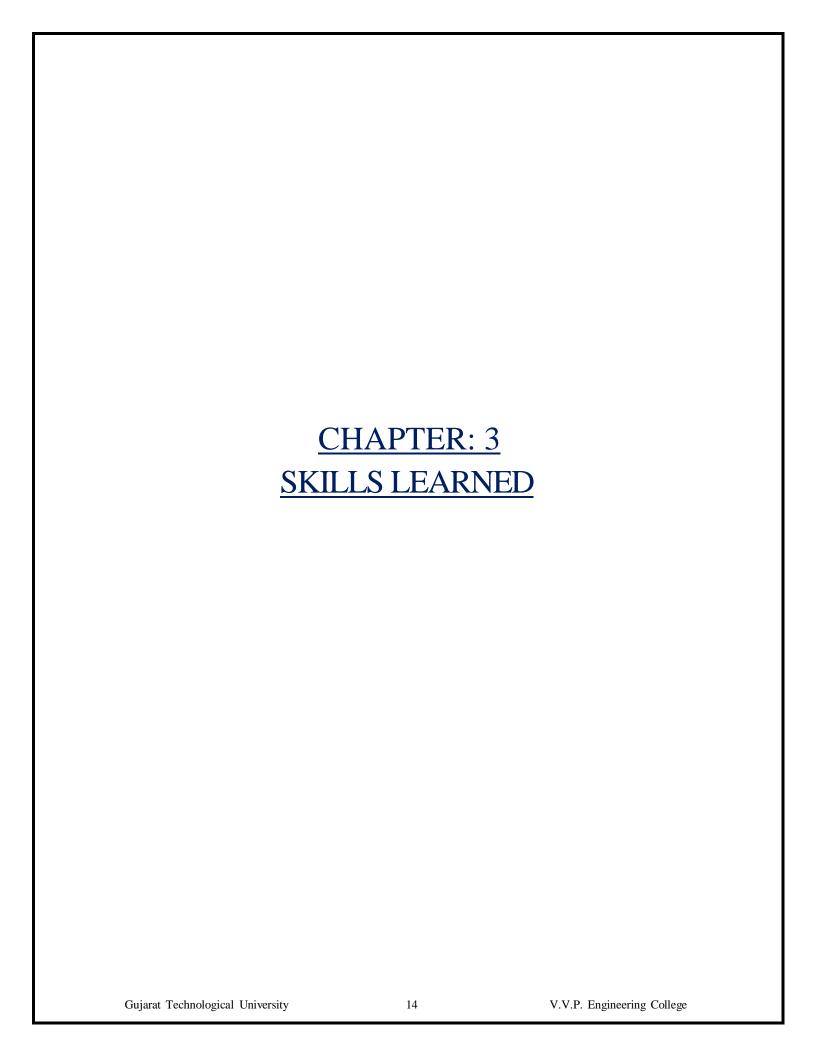


Fig: 2.4.1 Gain Chart



3.1About the Skill/s:

- > During the course of the internship, I acquired a range of technical and professional skills that are critical in the field of data science and digital innovation. The major skills learned include:
 - **Python Programming** Used as the core language for implementing data preprocessing, visualization, and machine learning algorithms.
 - NumPy & Pandas Applied for efficient handling of customer data, cleaning datasets, and performing transformations.
 - **Data Preprocessing** Learned techniques such as handling missing values, removing duplicates, and normalizing features to prepare telecom data for modeling.
 - Exploratory Data Analysis (EDA) Conducted statistical analysis and created visualizations (heatmaps, bar charts, histograms, scatter plots) to identify churn patterns.
 - Feature Engineering Designed new features like tenure and usage metrics to enhance model accuracy.
 - Machine Learning Algorithms Implemented Logistic Regression, Decision Trees,
 Random Forest, and Gradient Boosting to predict churn.
 - **Model Evaluation** Gained skills in applying accuracy, precision, recall, F1-score, and ROC-AUC to assess model performance.
 - **Data Visualization Tools** Used Matplotlib and Seaborn to represent patterns, customer segments, and prediction outcomes clearly.
 - **Problem-Solving with Analytics** Understood how business challenges (like churn) can be solved through data-driven models.
 - **Reporting & Documentation** Prepared structured reports, graphs, and results for communication with mentors and stakeholders.

3.2How do I learn the skill/s:

➤ The skills were learned through a combination of structured learning, hands-on projects, and real-time mentorship:

• Learning through Hands-on Practice

- Gained skills by directly working on telecom customer churn datasets.
- Practiced data preprocessing, cleaning, and feature engineering on real data.

• Step-by-Step Project Activities

- Each day of the internship was designed with practical tasks like using NumPy, Pandas,
 Matplotlib, Seaborn, and Scikit-learn.
- Completed mini-projects such as creating visualizations, handling missing values, and implementing algorithms, which gradually built my confidence.

Exploratory Data Analysis (EDA)

- Learned to analyze customer patterns, detect trends, and visualize relationships between features and churn.
- Strengthened my ability to interpret data-driven insights.

Model Development and Evaluation

- Gained hands-on experience in building machine learning models (Logistic Regression, Decision Trees, Random Forest, Gradient Boosting).
- Learned to evaluate models using accuracy, precision, recall, F1-score, and ROC-AUC.

Problem-Solving by Experimentation

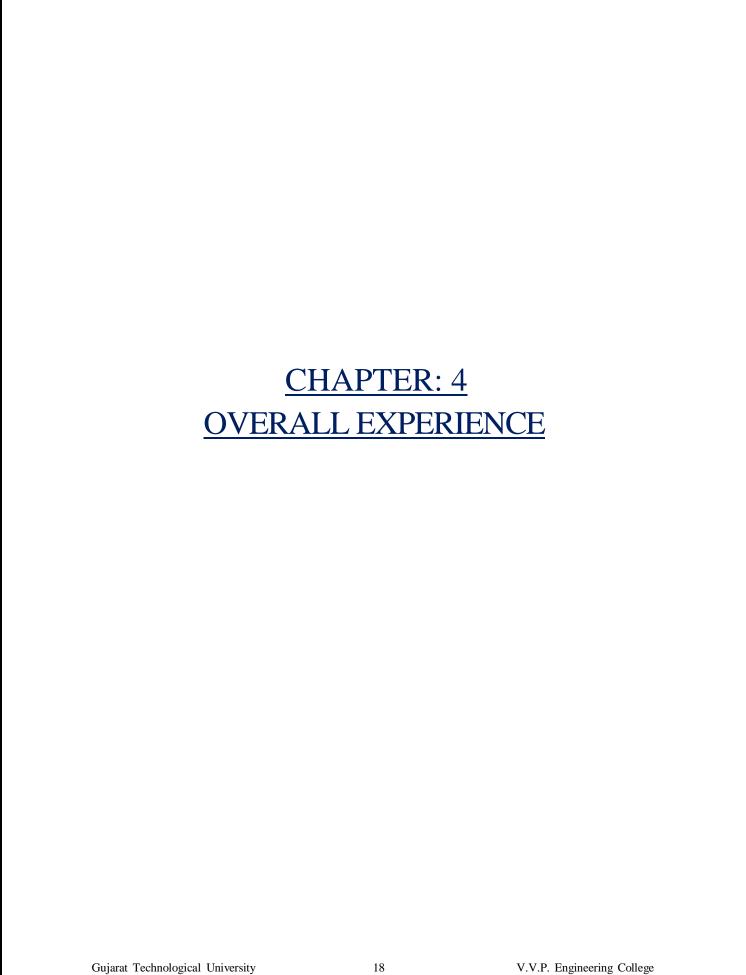
- Improved skills by experimenting with different models, hyperparameter tuning, and comparing performances.
- Understood how to troubleshoot errors and optimize models.

Documentation and Reporting

- Enhanced skills in documenting workflows, creating summaries, and presenting insights clearly.

• Independent & Guided Learning

- Followed mentor guidance while also self-learning through official documentation (Pandas, Scikit-learn, Kaggle resources).
- This combination helped me to develop both **technical expertise** and **self-learning skills**.

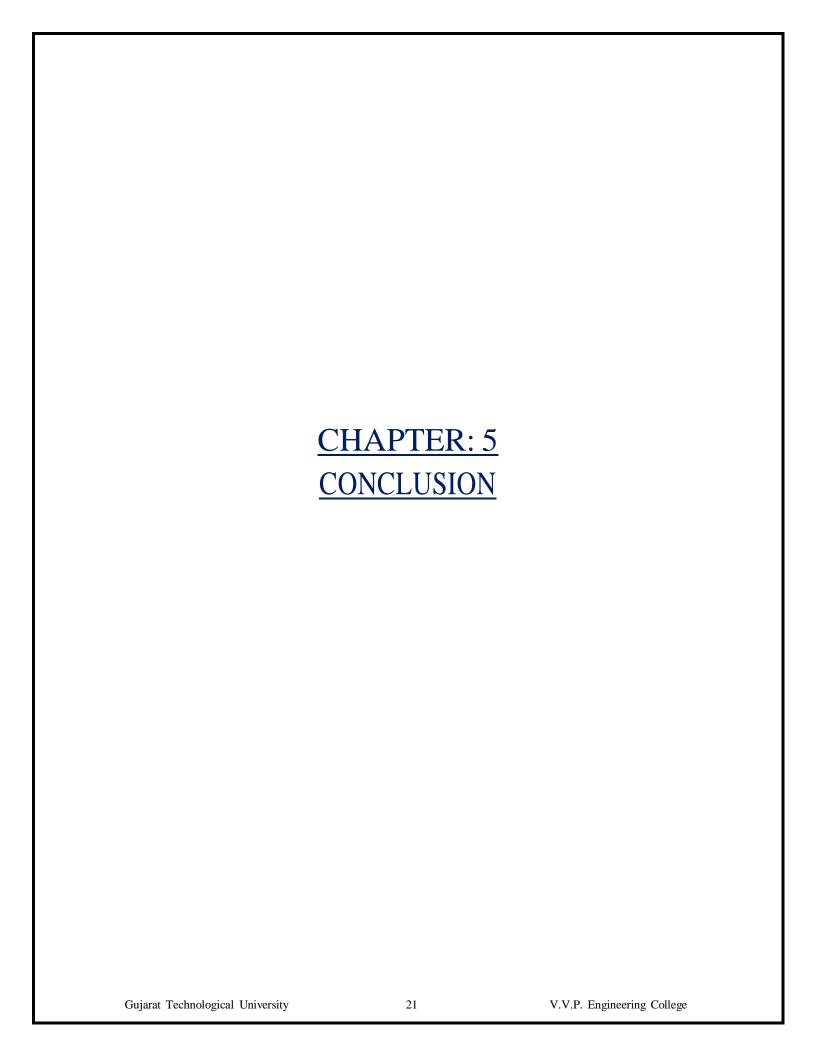


4.1Technical Experience:

- Gained hands-on experience in **Python programming** for data science and analytics.
- Worked extensively with NumPy and Pandas for data manipulation, cleaning, and preprocessing.
- Performed Exploratory Data Analysis (EDA) using Matplotlib and Seaborn for visualization.
- Implemented **data preprocessing techniques** such as handling missing values, removing duplicates, and outlier detection.
- Applied **feature engineering** and selection methods to improve model performance.
- Built and evaluated **machine learning models** (Logistic Regression, Decision Trees, Random Forest, Gradient Boosting).
- Gained exposure to model evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
- Practiced **hyperparameter tuning** using Grid Search and Random Search techniques.
- Developed understanding of **ensemble methods** for improving prediction accuracy.
- Learned to integrate predictive models with a **user-friendly interface** for practical usage.
- Documented workflows and insights for **reporting and presentation** purposes.

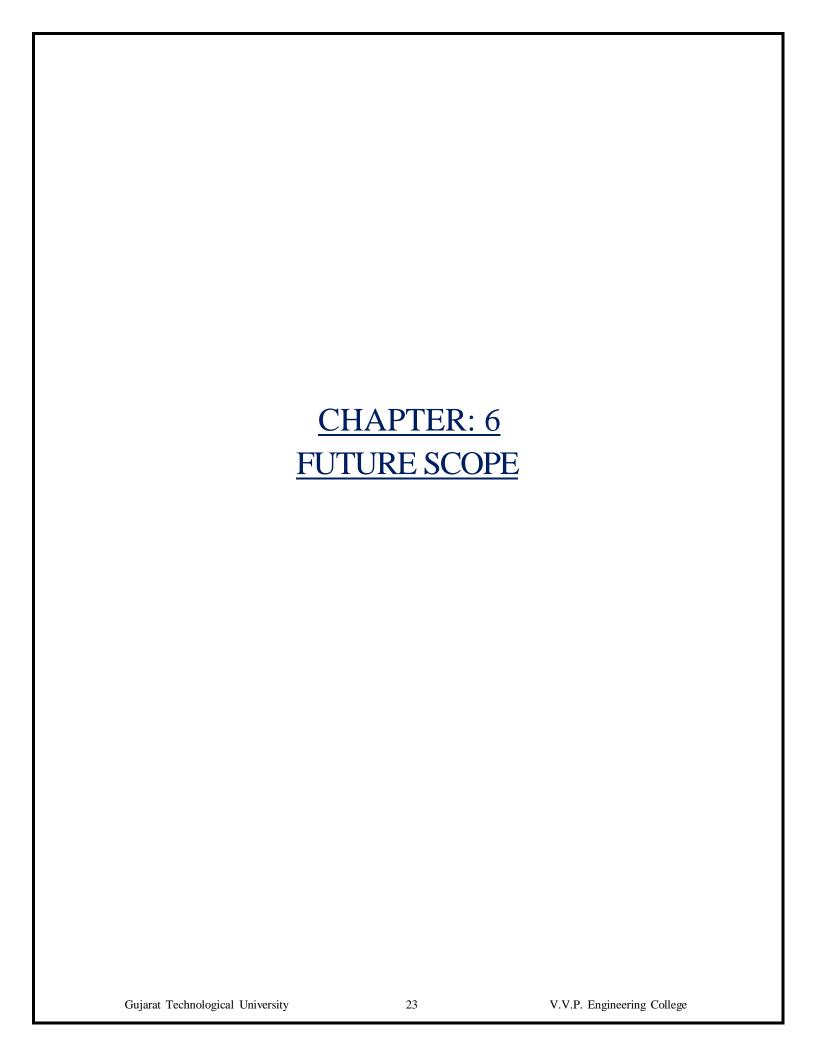
4.2Personal Experience:

- Beyond the technical learning, the internship experience was personally transformative and professionally enriching. It contributed to my overall growth in the following ways:
- **Confidence Building** Gained confidence in applying machine learning concepts to solve real-world business problems.
- **Time Management** Learned to organize daily tasks and complete activities within deadlines during the 30-day internship.
- Problem-Solving Skills Improved analytical thinking by handling missing data, noisy data, and feature selection challenges.
- **Technical Growth** Strengthened programming skills in Python, data visualization, and machine learning model development.
- **Communication Skills** Enhanced ability to explain technical results in a simple, business-oriented manner.
- **Self-Learning** Developed independent learning habits by exploring new tools, libraries, and datasets.
- Professional Exposure Understood how data-driven insights contribute to customer retention strategies in the telecom sector.



5.1Conclusion:

- The internship project on Telecom Customer Churn Analysis using Machine Learning successfully demonstrated how advanced data analytics can be applied to solve real-world business challenges. By analyzing customer demographics, service usage, and account information, the project identified significant factors contributing to customer churn. Multiple machine learning models such as Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting were implemented and evaluated to predict churn with improved accuracy.
- The developed system provides telecom providers with a decision-support tool that enables
 them to predict potential churners and take proactive measures to retain valuable customers.
 This not only enhances customer satisfaction but also ensures sustainable business growth.
- Through this project, I gained practical experience in data preprocessing, feature
 engineering, exploratory data analysis, model building, and performance evaluation,
 along with insights into how predictive analytics can be integrated into customer relationship
 management.
- In conclusion, the project highlights the importance of **data-driven strategies** in the telecom sector and sets the foundation for future improvements such as real-time churn prediction, automated retention strategies, and integration with advanced AI models.



6.1 Future Scope:

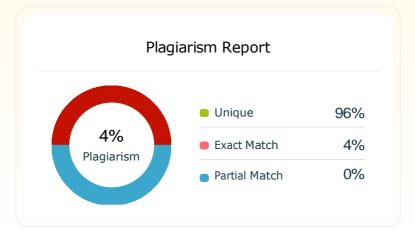
- **Integration with Real-Time Systems:** Deploy the churn prediction model in production with live customer data streams for real-time churn detection.
- **Inclusion of More Features:** Expand dataset with customer feedback, social media sentiment, and call center interactions to enhance prediction accuracy.
- **Automated Retention Strategies:** Link the model with CRM systems to automatically trigger personalized offers or retention campaigns.
- **Advanced Algorithms:** Apply deep learning models (e.g., LSTM, CNN) to capture complex sequential patterns in customer behavior.
- Scalability: Develop cloud-based churn prediction services for telecom operators handling millions of customers.
- Explainable AI (XAI): Implement model interpretability tools (e.g., SHAP, LIME) to provide clear insights into why a customer is predicted to churn.
- **Cross-Industry Adaptation:** Extend the churn analysis framework to other domains such as banking, e-commerce, and subscription-based businesses.

Team Id: 792975 Bibliography CHAPTER: 7 **BIBILOGRAPHY** 25

7.1 Bibliography:

- 1. Tom Fawcett, "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking", O'Reilly Media, 2013.
- 2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning", Springer, 2009.
- **3.** Géron, Aurélien, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow", O'Reilly Media, 2019.
- **4.** W. Verbeke, D. Martens, C. Mues, B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques", Expert Systems with Applications, 2011.
- 5. Scikit-learn Documentation Machine Learning in Python https://scikit-learn.org/stable/
- **6.** Pandas Documentation Data Analysis in Python https://pandas.pydata.org/
- 7. NumPy Documentation Scientific Computing with Python https://numpy.org/
- **8.** Matplotlib Documentation Data Visualization in Python https://matplotlib.org/stable/index.html
- 9. Seaborn Documentation Statistical Data Visualization https://seaborn.pydata.org/





Primary Sources



1 https://ibmorg-public.s3.us-e...



As a business that applies technology to a wide range of challenges, IBM also uses innovation and expertise to engage the SDG effort across all 17 goals.Missing: set explored objectives.

Excluded URL (s)



01 None

Content

Internship Report: IBM SkillsBuild Al-Powered Health Companion Project 2.1 Identi ed Challenges

During the course of the internship, the following critical problems were recognized and addressed:

Limited availability of healthcare facilities in isolated and marginalized communities.

Absence of tools for individuals to consistently monitor both their physical and mental health.

Lack of awareness and engagement in preventive healthcare practices among youth and students.

The need for early detection of health risks using advanced technologies such as AI.

Dif culty in simplifying technical Al concepts for individuals with no technical background.

These issues were approached through user research, ideation sessions, and the development of a user-centric AI health assistant in alignment with UN SDG 3: Good Health & Well-being.

2.2 My Role and Contributions

As an intern under IBM's SkillsBuild program, I undertook several responsibilities:

Research and Analysis

Explored the goals and targets of SDG 3 and studied current healthcare challenges globally.

Reviewed and assessed existing digital health tools to identify gaps and improvement areas.

Technical Skill Development

Completed training modules on Arti cial Intelligence, Natural Language Processing (NLP), Design Thinking, and Data Privacy.

Earned certi cations for foundational skills in both technical and professional domains.

Project Development

Helped outline the architecture and ow of the Al-driven wellness assistant.

Created initial design drafts and layouts using digital tools such as Figma.

Documentation and Reporting

Contributed to compiling reports and materials for nal presentation and submission.

2.3 Daily Work Summary

Day(s) Tasks / Activities

- 1-2 Orientation to the IBM SkillsBuild platform, set goals, and explored SDG 3 objectives.
- 3-4 Completed foundational courses in Al and the principles of Design Thinking.
- 5-6 Identi ed key healthcare challenges and de ned the scope of the project.
- 7-8 Designed feature logic for health monitoring and symptom tracking.
- 9-10 Integrated concepts for mental health support using NLP techniques.
- $11\mbox{-}12$ Finalized the prototype and evaluated ethical and privacy aspects.
- 13 Compiled documentation and re ected on project insights.
- 14 Presented and submitted the nal deliverable.

4.1 Technical Learnings

The internship equipped me with applied knowledge in several technological areas:

Arti cial Intelligence (AI):

Gained practical understanding of AI models, machine learning basics, and prediction systems. These were directly used to develop the AI health assistant.

Natural Language Processing (NLP):

Applied NLP methods to enable conversational interfaces that support mental health check-ins.

Design Thinking:

Used IBM's user-centric methodology to build a solution that truly met users' wellness needs.

Data Privacy & Ethics in Al:

Explored responsible Al development, with attention to user data protection and fairness.

IBM Tools:

Learned to use IBM Watson, SkillsBuild Labs, and various guided learning

platforms to rainforce technical knowledge.	
References	
Table of Contents	
Table of Contents	