

**Alzheimer Disease Prediction Model Project Report**

Submitted By:

Adwika Singh

AIML 6 Week Course

[singh.adwika11@gmail.com](mailto:singh.adwika11@gmail.com)

**ACKNOWLEDGEMENT**

I thank the National Institute of Electronics and Information Technology (NIELIT) Ropar, Executive Director, and all our faculty members for allowing us to undergo this project on AI and Machine Learning (ML). I express deep gratitude to our teachers **Mr. Sarwan Singh** and  **Ms.Suman**, Faculty at NIELIT Ropar, for their constant encouragement and guidance during this work and for always inspiring us. We express deep gratitude to them for mentoring the key concepts of the AI, ML, and Python language and skilled us to accomplish this work well today.

We also thank all other NIELIT Ropar Campus members for all the motivation and help they gave us during my project.

**Table of Contents**

1. Abstract

2. Chapter 1 – Introduction

1.1 Artificial Intelligence and Machine Learning

1.2 Applications of Machine Learning

1.3 Real World Examples of Machine Learning

1.4 About Python Programming Language

1.5 History of Python

1.6 Features of Python

1.7 Role of Python in Machine Learning

1.8 Jupyter Notebook

1.9 Google Colab

3. Chapter 2 – Training Work Undertaken

2.1 Project Introduction – Alzheimer's Disease Prediction

2.2 Libraries Used

2.3 Dataset

Patient Information

Demographic Details

Lifestyle Factors

Medical History

Clinical Measurements

Cognitive and Functional Assessments

Symptoms

Diagnosis Information

Confidential Information

2.4 Data Preprocessing

Handling Missing Values

Dropping Irrelevant Columns

Encoding Categorical Variables

Splitting the Data

Standardizing Features

2.5 Model Selection

2.6 Hyperparameter Tuning

2.7 Model Evaluation

2.8 Web Application Development

2.9 Deployment

4. Chapter 3 – Results

5. Chapter 4 – Conclusion and Future Scope

4.1 Conclusion

4.2 Future Scope

6. References

**Abstract**

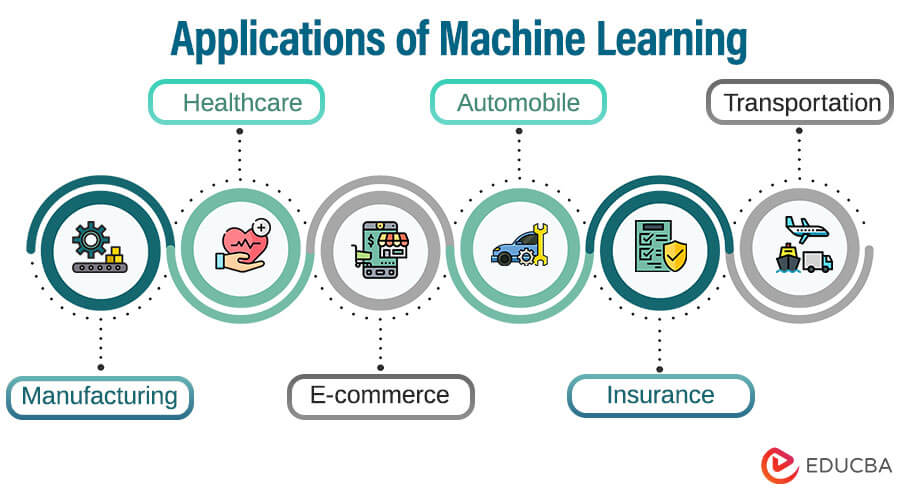
This project aims to predict Alzheimer's disease using machine learning techniques, specifically logistic regression. Alzheimer's disease is a progressive neurodegenerative disorder that leads to cognitive decline and memory loss. Early prediction of Alzheimer's can aid in better management and treatment of the disease. Using a comprehensive dataset containing demographic, lifestyle, medical history, and clinical measurement features, we developed a predictive model to classify patients as either having Alzheimer's disease or not. The dataset was obtained from Kaggle, and the model achieved an accuracy of 83% on the test set after hyperparameter tuning.

Chapter 1 – Introduction

**1.1 Artificial Intelligence and Machine Learning**

Artificial Intelligence (AI) is the field of computer science that focuses on creating systems capable of performing tasks that typically require human intelligence. Machine Learning (ML) is a subset of AI that involves training algorithms to learn from data and make predictions or decisions without explicit programming. ML has revolutionized many industries by providing powerful tools to analyze large datasets and extract meaningful insights.

**1.2 Applications of Machine Learning**



Machine learning has a wide range of applications across various fields:

1. Healthcare: Predicting diseases, personalized treatment plans, medical imaging analysis, and drug discovery.

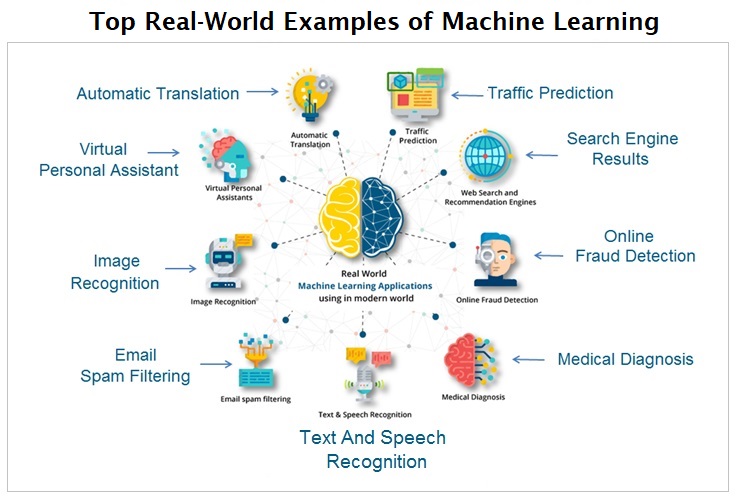
2. Finance: Fraud detection, stock market prediction, credit scoring, and algorithmic trading.

3. Retail: Recommendation systems, demand forecasting, and customer segmentation.

4. Transportation: Autonomous vehicles, route optimization, and traffic prediction.

5. Manufacturing: Predictive maintenance, quality control, and supply chain optimization.

**1.3 Real World Examples of Machine Learning**



1. Healthcare: IBM Watson for Oncology uses machine learning to provide evidence-based treatment options for cancer patients.

2. Finance: PayPal uses machine learning algorithms to detect and prevent fraudulent transactions.

3. Retail: Amazon's recommendation system suggests products to customers based on their browsing and purchase history.

4. Transportation: Tesla's Autopilot feature uses machine learning to enable semi-autonomous driving.

5. Manufacturing: General Electric uses machine learning for predictive maintenance of industrial equipment, reducing downtime and maintenance costs.

**1.4 About Python Programming Language**



Python is a high-level, interpreted programming language known for its simplicity and readability. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python's extensive libraries and frameworks make it a popular choice for various applications, including web development, data analysis, scientific computing, and machine learning.

**1.5 History of Python**

Python was created by Guido van Rossum and first released in 1991. The language was designed with an emphasis on code readability and simplicity. Python 2.0 was released in 2000, introducing features like list comprehensions and garbage collection. Python 3.0, released in 2008, brought significant changes to the language, including improvements to the standard library and better Unicode support. Today, Python is one of the most widely used programming languages in the world.

**1.6 Features of Python**

1. Simple and Easy to Learn: Python's syntax is straightforward, making it accessible for beginners.

2. Versatile: Python can be used for web development, data analysis, machine learning, automation, and more.

3. Extensive Libraries: Python has a rich ecosystem of libraries and frameworks, such as NumPy, pandas, scikit-learn, TensorFlow, and Django.

4. Community Support: Python has a large and active community, providing ample resources, tutorials, and third-party modules.

5. Cross-Platform: Python is available on various operating systems, including Windows, macOS, and Linux.

**1.7 Role of Python in Machine Learning**

Python's simplicity and powerful libraries make it an ideal language for machine learning. Libraries like scikit-learn, TensorFlow, Keras, and PyTorch provide tools for building, training, and evaluating machine learning models. Python's data manipulation libraries, such as pandas and NumPy, facilitate data preprocessing and analysis. Additionally, visualization libraries like Matplotlib and Seaborn enable effective data presentation.

**1.8 Jupyter Notebook**

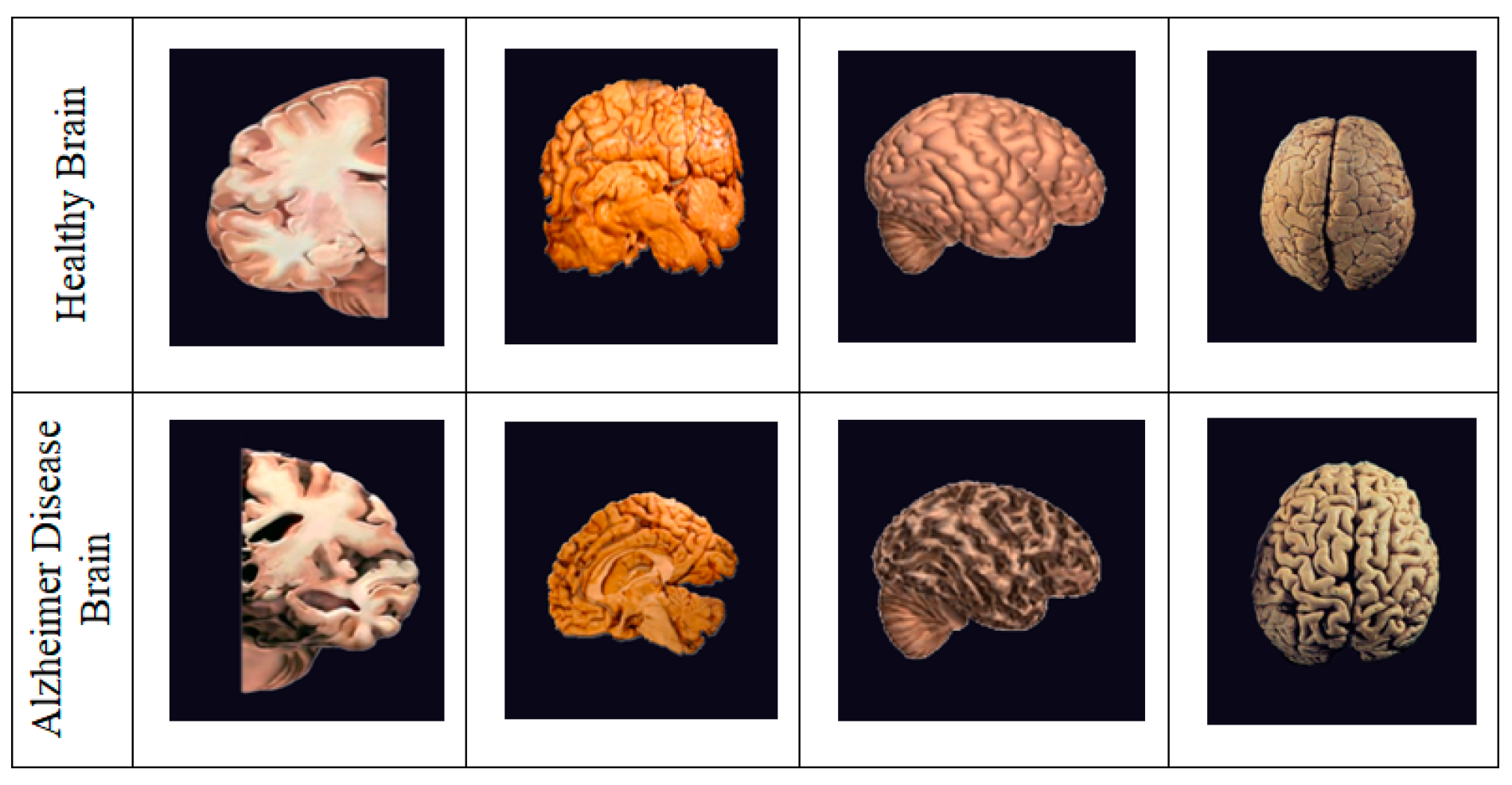
Jupyter Notebook is an open-source web application that allows users to create and share documents containing live code, equations, visualizations, and narrative text. It supports multiple programming languages, including Python, and is widely used in data science and machine learning for exploratory data analysis, visualization, and documentation.

**1.9 Google Colab**

Google Colab is a cloud-based platform that provides Jupyter Notebook environments with free access to computing resources, including GPUs and TPUs. It allows users to write and execute Python code in a web-based interface, making it convenient for machine learning projects that require substantial computational power.

**CHAPTER 2 – Training Work Undertaken**

**2.1 Project Introduction – Alzheimer's Disease Prediction**



Alzheimer's disease is a chronic neurodegenerative disease that affects millions of people worldwide. Early detection of Alzheimer's is crucial for managing the disease and improving the quality of life for patients. This project aims to develop a machine learning model to predict Alzheimer's disease based on a comprehensive dataset containing various patient attributes, including demographic details, lifestyle factors, medical history, and clinical measurements.

**2.2 Libraries Used**

The following libraries were used in this project:

1. pandas: For data manipulation and analysis.

2. numpy: For numerical computations.

3. scikit-learn: For machine learning model building and evaluation.

4. joblib: For saving and loading the trained model.

5. Flask: For building the web application.

**2.3 Dataset**



The dataset used in this project was obtained from Kaggle. It contains information on various patient attributes, including:

1. Patient Information

- PatientID: A unique identifier assigned to each patient.

2. Demographic Details

- Age: The age of the patients ranges from 60 to 90 years.

- Gender: Gender of the patients, where 0 represents Male and 1 represents Female.

- Ethnicity: The ethnicity of the patients, coded as follows:

- 0: Caucasian

- 1: African American

- 2: Asian

- 3: Other

- EducationLevel: The education level of the patients, coded as follows:

- 0: None

- 1: High School

- 2: Bachelor's

- 3: Higher

3. Lifestyle Factors

- BMI: Body Mass Index of the patients, ranging from 15 to 40.

- Smoking: Smoking status, where 0 indicates No and 1 indicates Yes.

- AlcoholConsumption: Weekly alcohol consumption in units, ranging from 0 to 20.

- PhysicalActivity: Weekly physical activity in hours, ranging from 0 to 10.

- DietQuality: Diet quality score, ranging from 0 to 10.

- SleepQuality: Sleep quality score, ranging from 4 to 10.

4. Medical History

- FamilyHistoryAlzheimers: Family history of Alzheimer's Disease, where 0 indicates No and 1 indicates Yes.

- CardiovascularDisease: Presence of cardiovascular disease, where 0 indicates No and 1 indicates Yes.

- Diabetes: Presence of diabetes, where 0 indicates No and 1 indicates Yes.

- Depression: Presence of depression, where 0 indicates No and 1 indicates Yes.

- HeadInjury: History of head injury, where 0 indicates No and 1 indicates Yes.

- Hypertension: Presence of hypertension, where 0 indicates No and 1 indicates Yes.

5. Clinical Measurements

- SystolicBP: Systolic blood pressure, ranging from 90 to 180 mmHg.

- DiastolicBP: Diastolic blood pressure, ranging from 60 to 120 mmHg.

- CholesterolTotal: Total cholesterol levels, ranging from 150 to 300 mg/dL.

- CholesterolLDL: Low-density lipoprotein cholesterol levels, ranging from 50 to 200 mg/dL.

- CholesterolHDL: High-density lipoprotein cholesterol levels, ranging from 20 to 100 mg/dL.

- CholesterolTriglycerides: Triglycerides levels, ranging from 50 to 400 mg/dL.

6. Cognitive and Functional Assessments

- MMSE: Mini-Mental State Examination score, ranging from 0 to 30. Lower scores indicate cognitive impairment.

- FunctionalAssessment: Functional assessment score, ranging from 0 to 10. Lower scores indicate greater impairment.

- MemoryComplaints: Presence of memory complaints, where 0 indicates No and 1 indicates Yes.

- BehavioralProblems: Presence of behavioral problems, where 0 indicates No and 1 indicates Yes.

- ADL: Activities of Daily Living score, ranging from 0 to 10. Lower scores indicate greater impairment.

7. Symptoms

- Confusion: Presence of confusion, where 0 indicates No and 1 indicates Yes.

- Disorientation: Presence of disorientation, where 0 indicates No and 1 indicates Yes.

- PersonalityChanges: Presence of personality changes, where 0 indicates No and 1 indicates Yes.

- DifficultyCompletingTasks: Presence of difficulty completing tasks, where 0 indicates No and 1 indicates Yes.

- Forgetfulness: Presence of forgetfulness, where 0 indicates No and 1 indicates Yes.

8. Diagnosis Information

- Diagnosis: Diagnosis status for Alzheimer's Disease, where 0 indicates No and 1 indicates Yes.

9. Confidential Information

- DoctorInCharge: This column contains the name of the doctor in charge.

**2.4 Data Preprocessing**

The data preprocessing steps involved the following:

1. Handling Missing Values: Any missing values in the dataset were handled appropriately. If necessary, imputation techniques were used to fill missing values based on the mean or median of the respective column.

2. Dropping Irrelevant Columns: Columns that were not relevant to the prediction task, such as 'PatientID' and 'DoctorInCharge', were dropped from the dataset.

3. Encoding Categorical Variables: Categorical variables such as 'Gender', 'Ethnicity', and 'EducationLevel' were encoded using numerical values.

4. Splitting the Data: The dataset was split into training and test sets in an 80:20 ratio.

5. Standardizing Features: The features were standardized to have a mean of 0 and a standard deviation of 1.

**2.5 Model Selection**

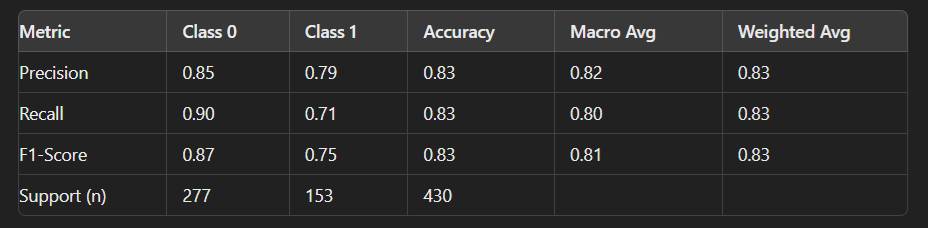
Logistic Regression was chosen as the model for predicting Alzheimer's disease. Logistic regression is a simple yet powerful classification algorithm that is suitable for binary classification tasks.

**2.6 Hyperparameter Tuning**

Hyperparameter tuning was performed using Grid Search to find the best parameters for the logistic regression model. The best parameter found was 'C': 1.

**2.7 Model Evaluation**

The model was evaluated using the test set, and the following classification report was generated:



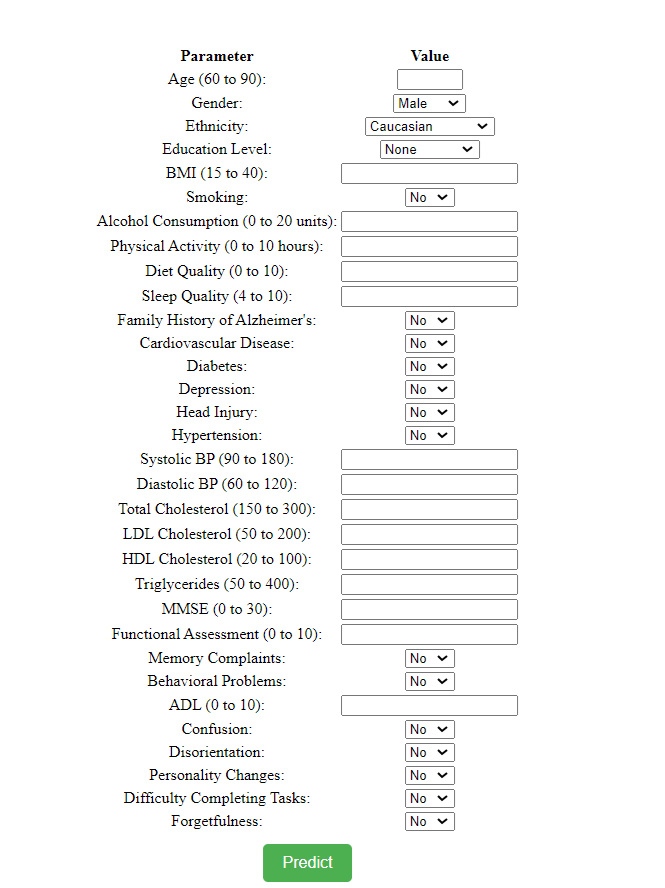
The model achieved an accuracy of 83% on the test set.

**2.8 Web Application Development**

A web application was developed using Flask to provide a user-friendly interface for predicting Alzheimer's disease. The application allows users to input patient details and get a prediction on whether the patient has Alzheimer's disease or not.

**2.9 Deployment**

The web application was deployed on Glitch, a platform for building and sharing web applications. The deployment process involved creating a new project on Glitch, uploading the necessary files, and configuring the environment to run the Flask application.



**Conclusion**

This project successfully developed a machine learning model to predict Alzheimer's disease with an accuracy of 83%. The model was deployed as a web application, providing a user-friendly interface for making predictions. This project demonstrates the potential of machine learning in healthcare and the importance of early detection of Alzheimer's disease. Future work could involve using more advanced machine learning techniques and incorporating additional features to improve the model's accuracy.

**References**

* Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. CreateSpace.
* Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
* Chollet, F. (2015). Keras. GitHub. Retrieved from <https://github.com/fchollet/keras>
* Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95.
* Oliphant, T. E. (2006). A guide to NumPy (Vol. 1, p. 85). USA: Trelgol Publishing.