

ALZHEIMER'S PREDICTION

ASHOK VENKATESH 100003522
ACHAL AJITH 100004415
PRAFUL PRAKAS 10000



TABLE OF CONTENTS

01.

Business
Understanding

02.

Data Understanding

03.

Data Preparation

04.

Modeling

TABLE OF CONTENTS

05.

Evaluation

06.

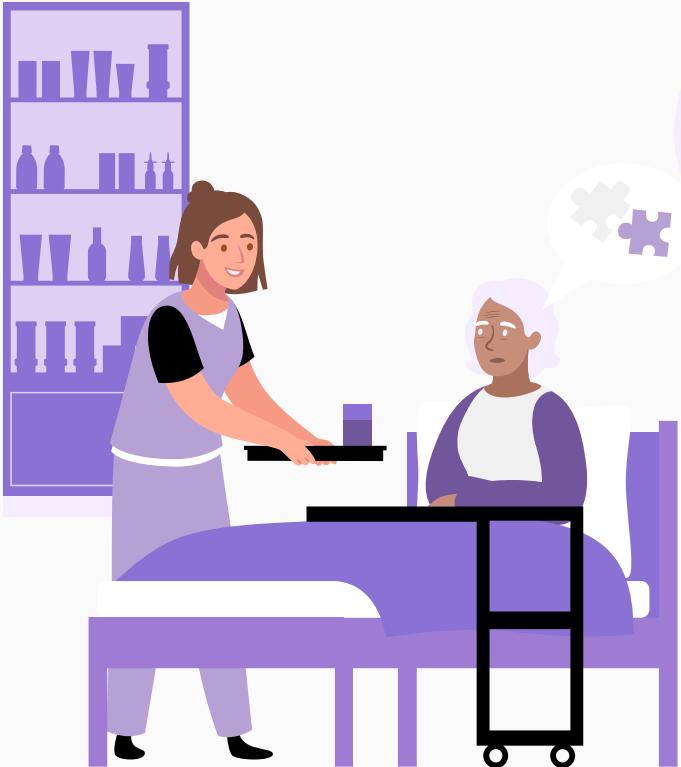
Deployment

07.

Conclusion

01.

Business Understanding



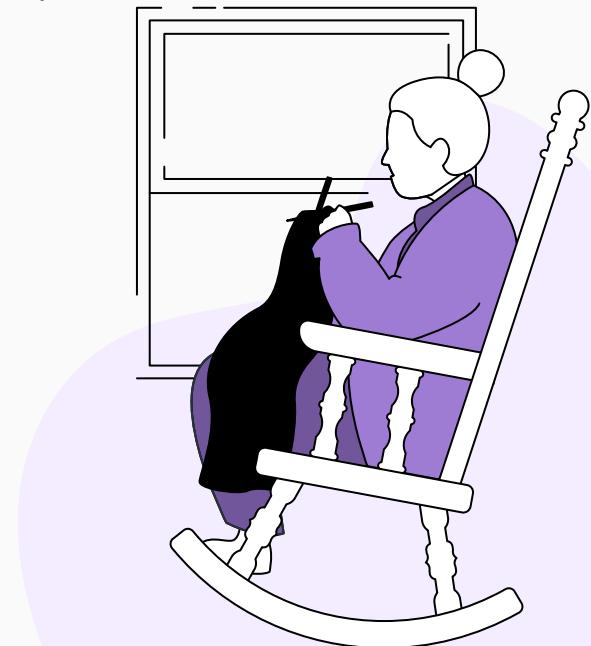
WHY DO WE NEED IT?

Business Problem

- Alzheimer's disease causes progressive cognitive decline and loss of independence
- Traditional diagnosis methods are:
 - Expensive (MRI, PET scans)
 - Time-consuming
 - Dependent on specialist availability
- Delayed diagnosis reduces treatment effectiveness

Business Value

- Early detection enables timely intervention
- Cost-effective screening compared to imaging tests
- Scalable solution for hospitals and telemedicine platforms



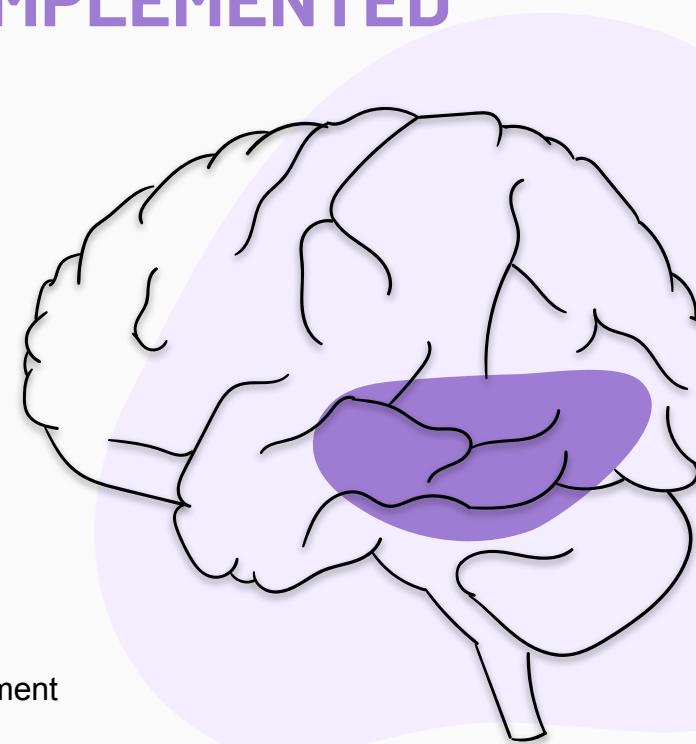
CHALLENGES AND SOLUTIONS IMPLEMENTED

Challenges

- Class imbalance between Alzheimer's and non-Alzheimer's cases
- Identifying the most relevant clinical features
- Limited interpretability of complex models (XGBoost)
- Risk of false negatives and ethical concerns in healthcare predictions

Solutions & Future Scope

- Used XGBoost for robust and accurate predictions
- Evaluated model using recall and confusion matrix, not just accuracy
- Positioned model as a decision-support tool, not a diagnostic replacement
- **Unsolved challenges:** need for clinical validation, bias reduction, and improved model explainability



02.

Data Understanding



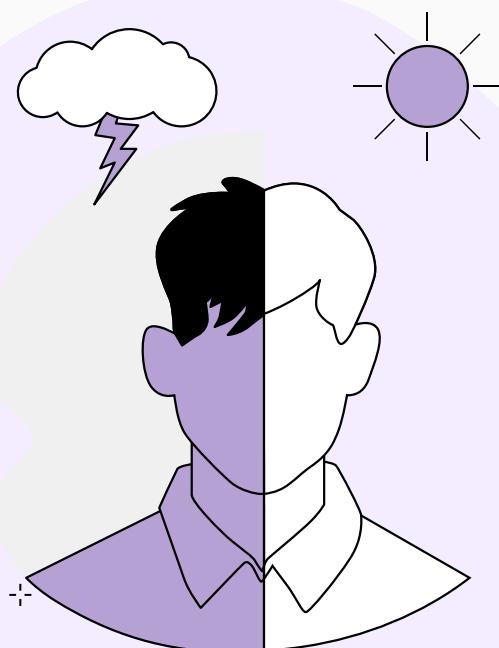
◦ DATASET AND CHALLENGES

Data Description

- Dataset contains demographic and clinical features (Dataset source: <https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset>)
- Key columns:
 - Age, Gender, EducationLevel
 - Cognitive scores: MMSE, CDR, MemoryScore
- Target variable: Diagnosis (Alzheimer's / Non-Alzheimer's)

Challenges Identified

- Missing values in MMSE and cognitive score fields
- Mixed data types (Gender categorical, scores numerical)
- Class imbalance in Diagnosis
- No clinical metadata explaining feature collection



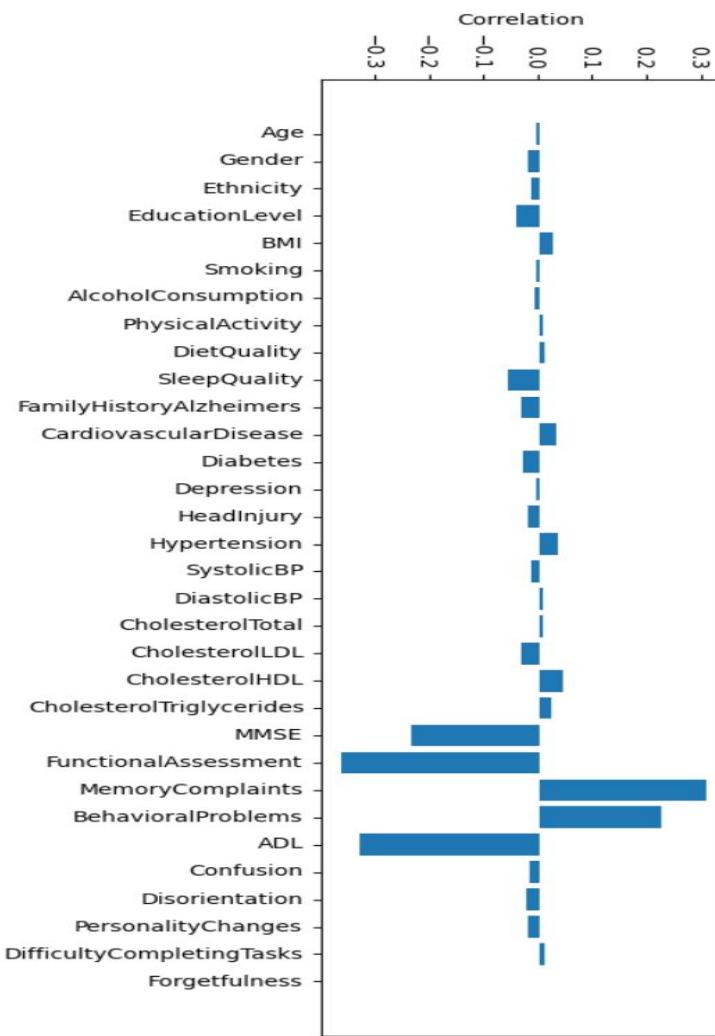
DATA INSIGHTS AND LIMITATIONS

Key Data Insights

- Pearson correlation analysis shows:
 - MMSE, FunctionalAssessment, and ADL are negatively correlated with Diagnosis
 - MemoryComplaints and BehavioralProblems are positively correlated with Diagnosis

Solutions & Open Issues

- Applied data cleaning and encoding for categorical variables
- Used correlation analysis to guide feature selection
- Unsolved challenges: demographic bias, limited real-world validation, and dependency on dataset quality





03.

Data Preparation

CHALLENGES AND SOLUTIONS IMPLEMENTED

Feature Engineering

- Removed irrelevant attributes to reduce noise and missing diagnostic records
- Encoded categorical diagnosis labels using Label Encoding
- Prepared dataset optimized for tree-based models (XGBoost)

Challenges & Solutions

- Missing clinical values → Applied median imputation
- Limited feature availability → Adapted model to use available, medically relevant features
- Non-numeric diagnosis labels → Converted using label encoding
- Risk of overfitting → Used feature selection and train–test split



04.

Modeling



MODEL SELECTION

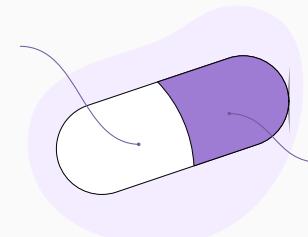
Models Evaluated

- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- XGBoost Classifier

Model Selection Rationale

- Models were compared using accuracy, recall, and confusion matrix
- Recall was prioritized to minimize missed Alzheimer's cases
- XGBoost demonstrated:
 - Higher recall
 - Better class separation in the confusion matrix
 - Strong overall accuracy

MODEL	ACCURACY	RECALL(0)	RECALL(1)	CONFUSION MATRIX
RANDOM FOREST	94.42%	0.97	0.89	[270, 8] [16,136]
XG BOOST	95.35%	0.98	0.91	[271, 6] [14,139]
LOGISTIC REGRESSION	83.02%	0.89	0.73	[246,31] [42 ,111]
SVM	83.26%	0.90	0.71	[250,27] [45,108]





05. **Evaluation**

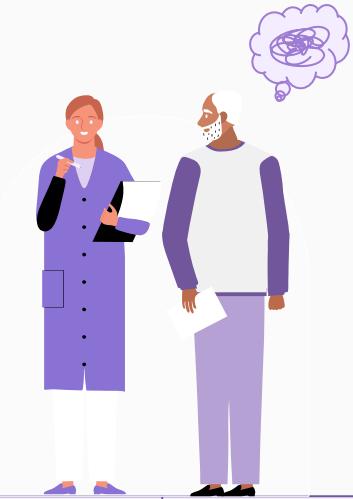
INTERPRETATION AND CHALLENGES

Performance Interpretation

- Recall of 91% shows the model successfully identifies most Alzheimer's cases
- Low false negatives (14 cases) reduces missed diagnoses
- Balanced precision and recall ensure reliable predictions

Challenges & Solutions

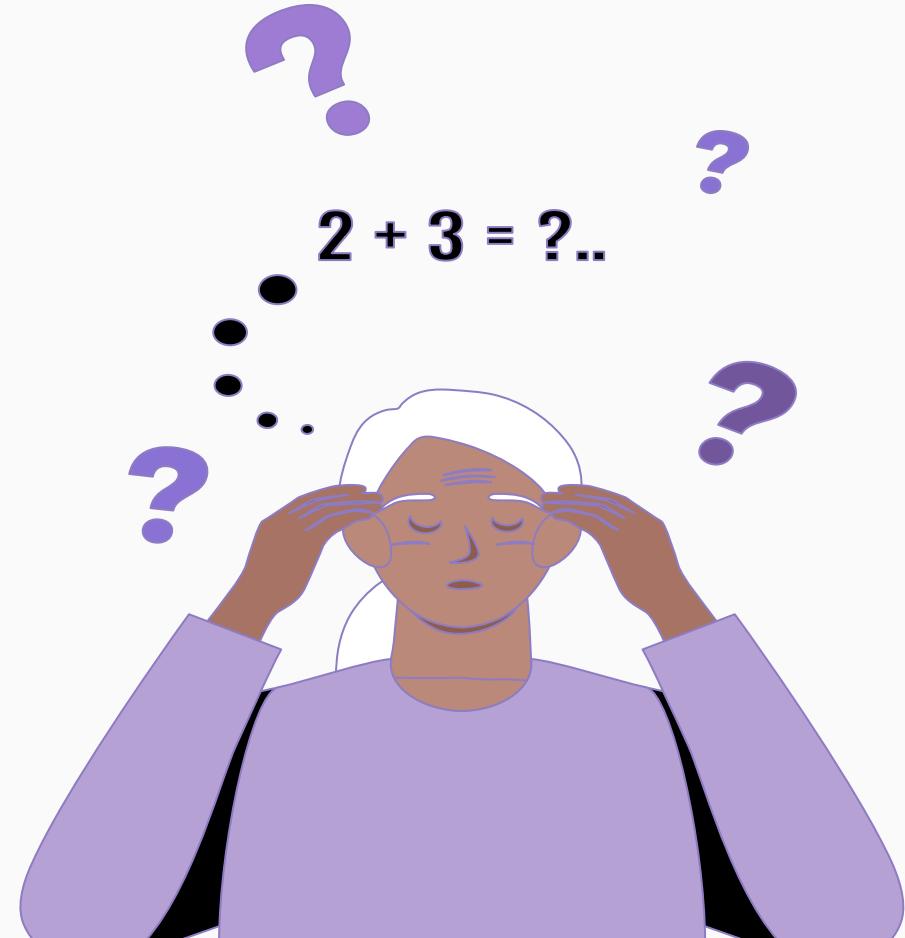
- Class imbalance in dataset → Used stratified 80/20 train–test split
- Risk of false negatives in medical prediction → Prioritized recall over accuracy
- Metric misinterpretation risk → Used confusion matrix and F1-score for validation



TN: 271	FP: 6
FN: 14	TP: 139

06.

Deployment



○ DEPLOYMENT- JUPYTER NOTEBOOK



Deployment Approach

- Trained XGBoost Alzheimer's prediction model
- Enabled deployment through an interactive Python-based prediction system
- Binary classification:
 - Alzheimer's
 - Not Alzheimer's

Deployment Challenges & Solutions

- Consistent feature ordering → Stored feature list with model
- Handling missing input values → Median-based preprocessing
- Model reuse across environments → Serialized model using Joblib
- User interpretation of results → Clear label mapping and confidence score

INPUT

```
===== INPUT SCALE =====
MMSE (Mini-Mental State Examination): 0-30
Functional Assessment: 0-5
Memory Complaints: 0 = No, 1 = Yes
Behavioral Problems: 0 = No, 1 = Yes
ADL (Activities of Daily Living): 0-6
=====
```

Enter the following values:

MMSE score: 23

Functional Assessment score: 4

Memory Complaints (0 = No, 1 = Yes): ↴ for history. Search history with c-↑/c-↓

OUTPUT

===== INPUT SCALE =====

MMSE (Mini-Mental State Examination): 0-30
Functional Assessment: 0-5 or according to your dataset scale
Memory Complaints: 0 = No, 1 = Yes
Behavioral Problems: 0 = No, 1 = Yes
ADL (Activities of Daily Living): 0-6 or according to your dataset scale
=====

Enter the following values:

MMSE score: 23

Functional Assessment score: 3

Memory Complaints (0 = No, 1 = Yes): 1

Behavioral Problems (0 = No, 1 = Yes): 0

ADL score: 2

Prediction: Alzheimer's

Confidence: 1.00

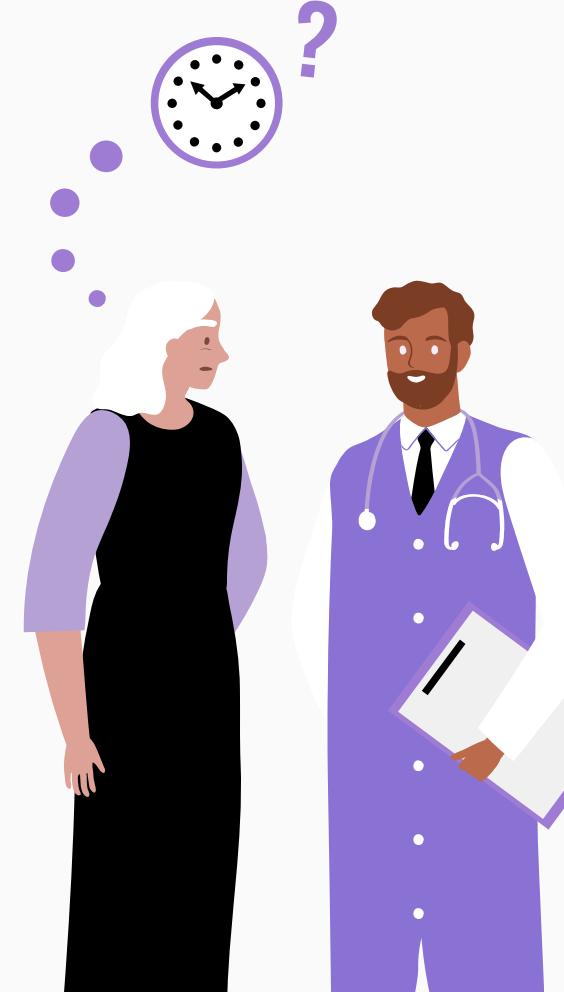


07.

Conclusion

CONCLUSION

- Successfully applied XGBOOST to create an Alzheimer's Prediction model with 95% accuracy and 91% recall effectively minimizing missed diagnoses.
- Key features such as MMSE, Functional Assessment, Memory Complaints, Behavioral Problems, and ADL were used to capture cognitive and functional impairment patterns.
- The trained model was deployed as an interactive prediction tool, providing real-time risk assessment with confidence scores.
- Suitable as a decision-support tool, not a diagnostic replacement



THANK YOU

