DATA SCIENCE: EXPLAINED & APPLIED

BY: MONTASER AYASRAH

001

DATA SCIENCE: EXPLAINED & APPLIED

HELLO!

Today's Topics

- Introduction to Data Science
- Key Principles of Data Science
- Application of Data Science
- Benefits of Using Data Science in Healthcare
- Data Science: From Research to Real-world Use
- The Power of Data Science in Healthcare
- Next Steps and Future Applications

DATA SCIENCE

Brief Introduction

Data science is a multidisciplinary field that uses techniques from statistics, computer science, and domain expertise to extract insights from data. It has applications in healthcare, finance, marketing, and countless other fields.

In healthcare, data science is revolutionizing the way we diagnose and treat diseases by analyzing complex medical data, like MRI images, to detect patterns that aren't always visible to the human eye.

BETTER UNDERSTADING

Things are often not fully understood until an example is given. To fully understand the applications and explanations of Data Science we will have an example project.

The project is a MRI Image Dementia Multi-classifier

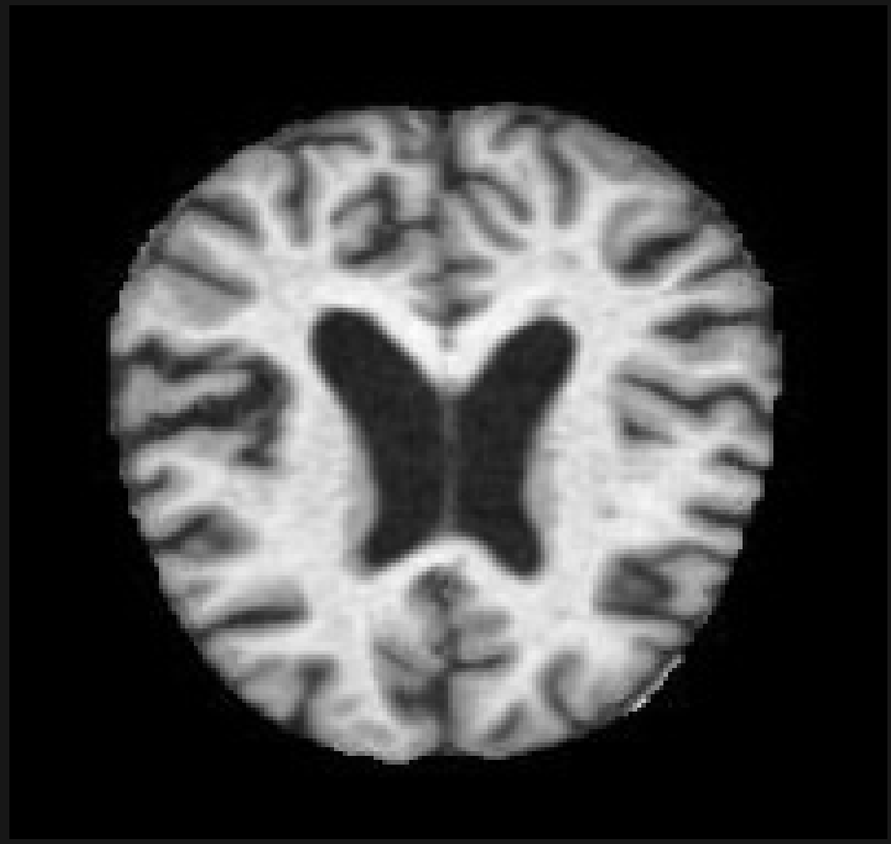
Big words?!?

We will explain further, buckle-up!

KEY PRINCIPLES OF DATA SCIENCE

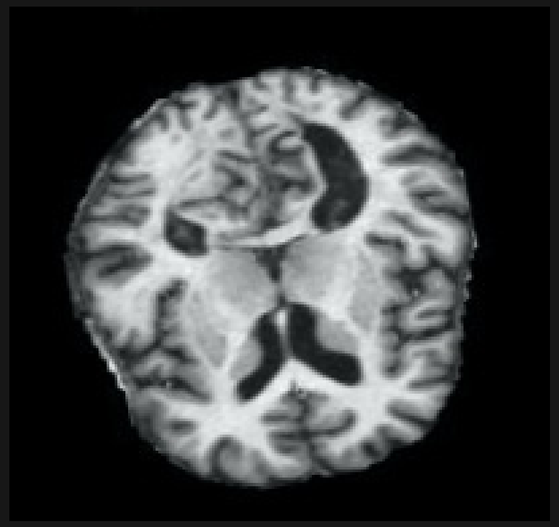
DATA COLLECTION

DATA SCIENCE STARTS WITH COLLECTING RELEVANT DATA. IN THIS PROJECT, THE DATA CONSISTS OF MRI SCANS OF THE BRAIN, WHICH CONTAIN VALUABLE INFORMATION ABOUT BRAIN STRUCTURE.

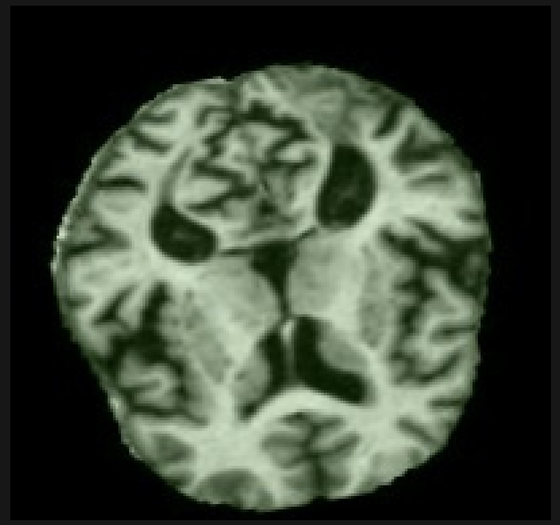


EXAMPLES:

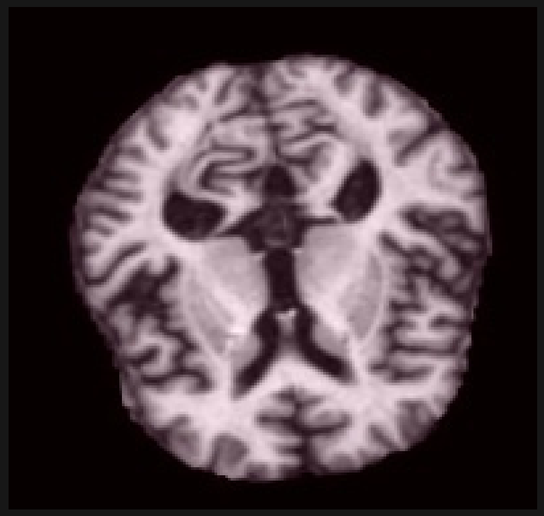
NON-DEMENTED



MILD DEMENTED



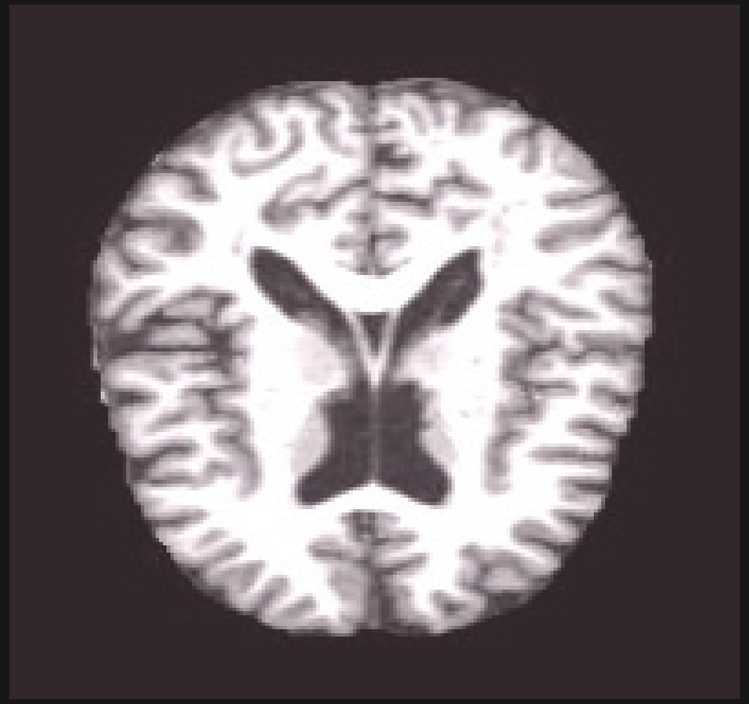
MODERATE DEMENTED



WHY IT MATTERS?
MRI SCANS CAPTURE FINE DETAILS OF BRAIN TISSUE. BY ANALYZING THESE DETAILS, DATA SCIENTISTS CAN DETECT CHANGES THAT MAY INDICATE DEMENTIA.

DATA PREPROCESSING

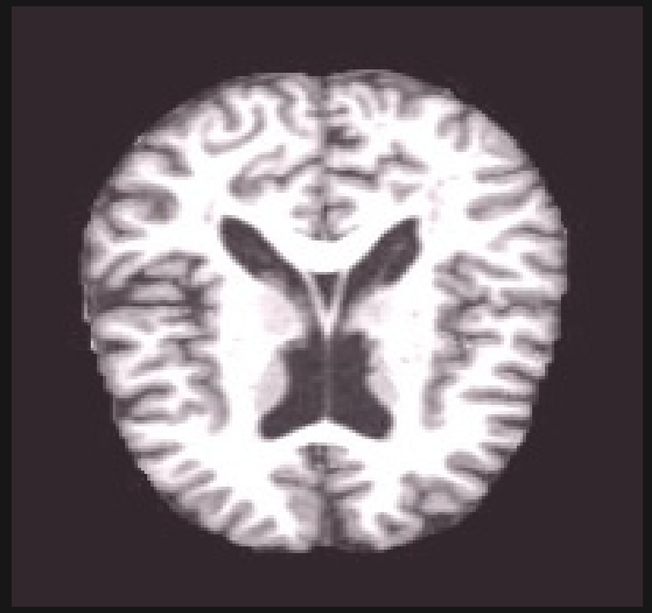
BEFORE THE RAW DATA CAN BE USED FOR ANALYSIS, IT NEEDS TO BE CLEANED AND TRANSFORMED INTO A FORMAT THAT MACHINE LEARNING ALGORITHMS CAN UNDERSTAND. THIS STEP IS CRUCIAL IN ANY DATA SCIENCE PROJECT.



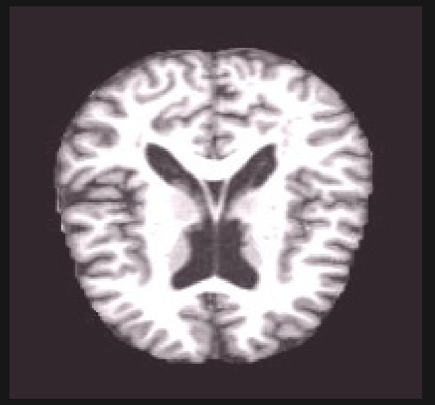
PREPROCESSING STEPS:

RESIZING: MRI SCANS COME IN VARIOUS SIZES. TO STANDARDIZE THEM, WE RESIZE ALL IMAGES TO 512X512 PIXELS, SO THE MACHINE LEARNING MODEL CAN TREAT EACH IMAGE UNIFORMLY.

EXAMPLE:



BEFORE

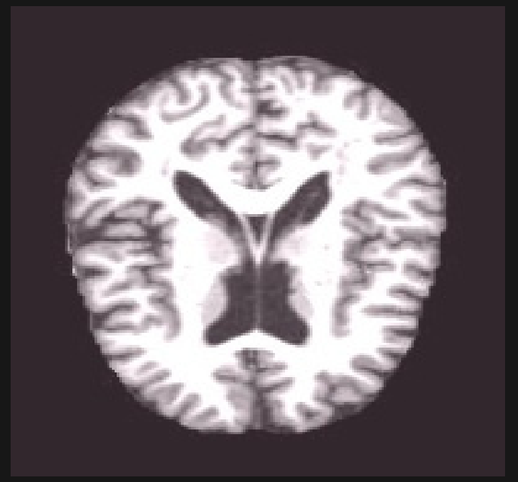


AFTER

PREPROCESSING STEPS:

GRAY-SCALING: MRI IMAGES ARE CONVERTED TO GRAYSCALE TO SIMPLIFY THE DATA WITHOUT LOSING IMPORTANT STRUCTURAL INFORMATION. SINCE COLOR ISN'T NECESSARY FOR DETECTING DEMENTIA, THIS STEP REDUCES THE COMPLEXITY OF THE IMAGES.

EXAMPLE:



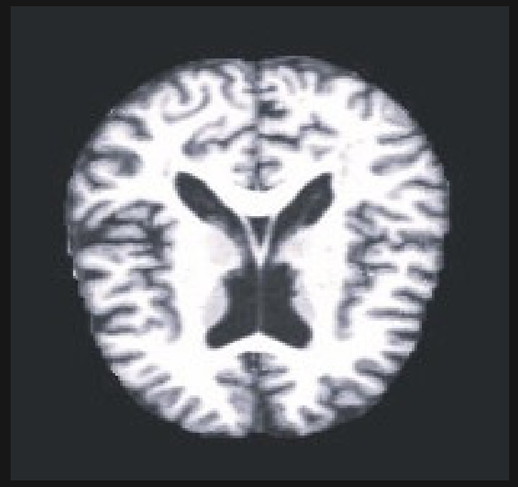
BEFORE

AFTER

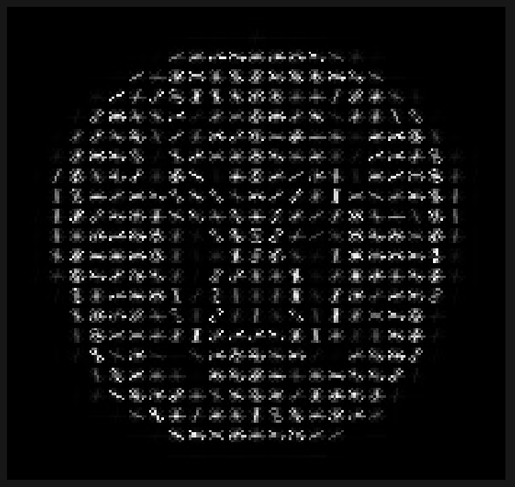
PREPROCESSING STEPS:

FEATURE EXTRACTION (HOG):
WE EXTRACT ESSENTIAL PATTERNS FROM THE MRI IMAGES USING THE HISTOGRAM OF ORIENTED GRADIENTS (HOG) METHOD, WHICH CAPTURES THE TEXTURE AND SHAPE OF OBJECTS IN THE IMAGE. THIS HELPS THE MODEL "SEE" THE PARTS OF THE BRAIN THAT MATTER MOST FOR DETECTING DEMENTIA.

EXAMPLE:



BEFORE



AFTER

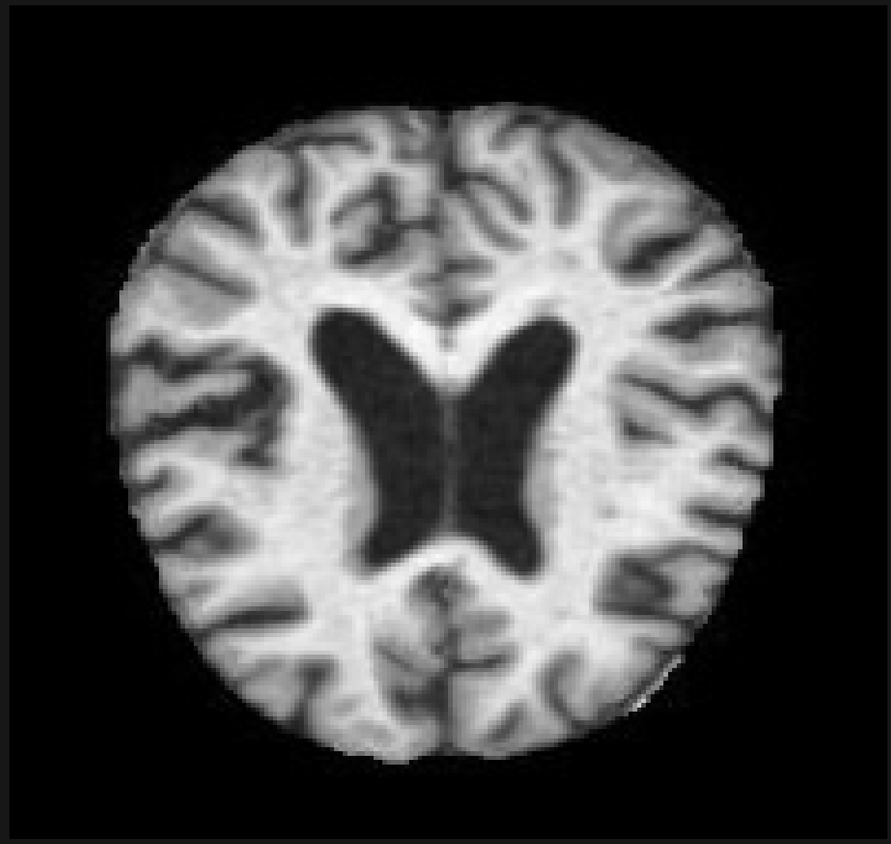
WHY PREPROCESS THO?

Preprocessing in data science is essential for transforming raw, inconsistent data into a clean, usable format that machine learning models can effectively interpret. It addresses issues like:

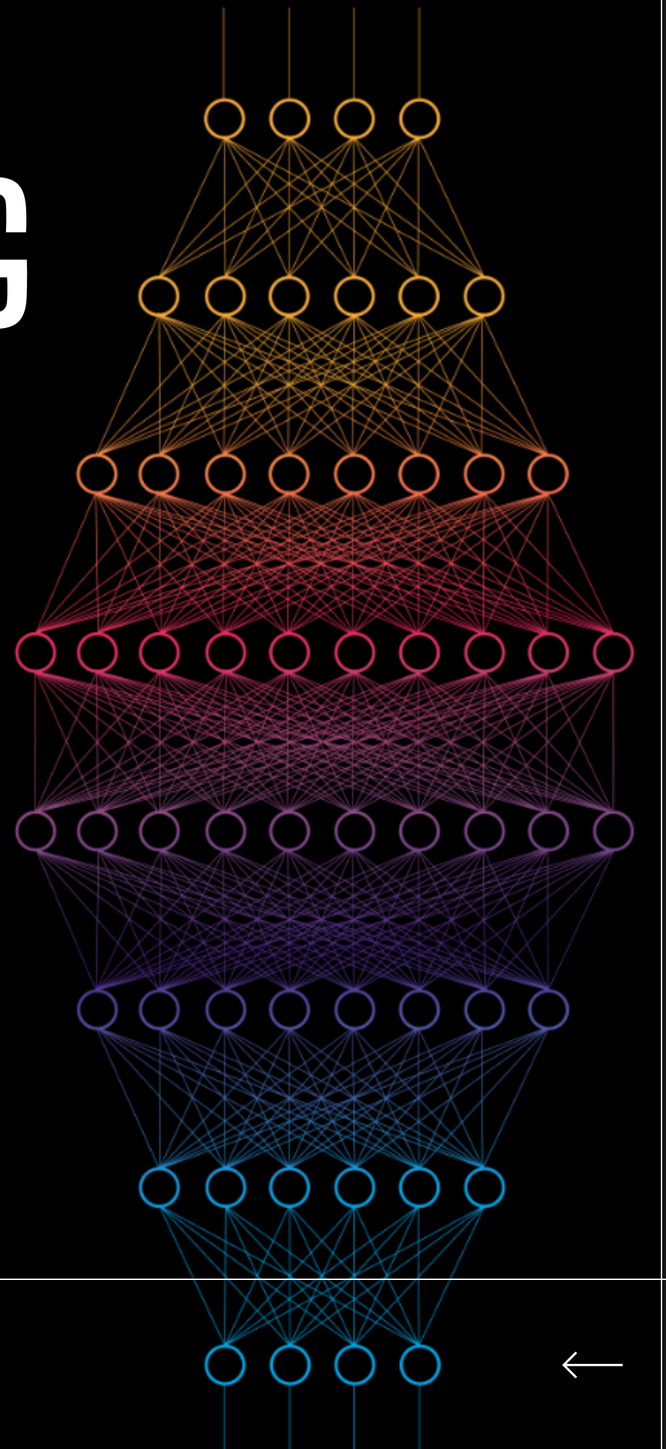
- Inconsistencies (e.g., different formats, resolutions),
- Noise reduction (removing irrelevant information),
- Missing data handling (imputing or removing incomplete entries),
- Data transformation (converting non-numeric data into machine-readable formats),
- Reducing computational complexity (simplifying data without losing important information).

DATA COLLECTION

DATA SCIENCE STARTS WITH COLLECTING RELEVANT DATA. IN THIS PROJECT, THE DATA CONSISTS OF MRI SCANS OF THE BRAIN, WHICH CONTAIN VALUABLE INFORMATION ABOUT BRAIN STRUCTURE.



MACHINE LEARNING



In data science, machine learning is the engine that drives predictions. Machine learning models "learn" from data to make predictions about new, unseen data.

MODELS USED IN THE PROJECT:

Random Forest:
A model that creates multiple decision trees and combines their results. It’s like asking multiple experts for their opinions and then combining them to make the best decision.

XGBoost:
A more advanced version of decision trees, optimized to handle large datasets like MRI images. It works well when you have complex data with subtle patterns.

TRAINING AND TESTING THE MODELS:

TO ENSURE THE MODELS ARE ACCURATE, WE SPLIT THE DATA INTO TWO PARTS:

Training Data: Used to teach the model how to recognize patterns in MRI images.

Testing Data: Used to check if the model can correctly classify new images.

WE TRAIN THE MODEL USING THE TRAINING DATA AND THEN TEST ITS ACCURACY USING THE TESTING DATA.

MODEL OPTIMIZATION:

ONCE A MODEL IS TRAINED, WE TUNE ITS PARAMETERS TO MAKE IT MORE ACCURATE. THIS PROCESS IS KNOWN AS HYPERPARAMETER OPTIMIZATION. IT’S LIKE FINE-TUNING A RECIPE BY ADJUSTING THE INGREDIENTS TO GET THE BEST FLAVOR.

IN THIS PROJECT, WE OPTIMIZED THE XGBOOST MODEL TO IMPROVE ITS ABILITY TO CORRECTLY CLASSIFY DIFFERENT STAGES OF DEMENTIA.

APPLICATIONS OF DATA SCIENCE IN MEDICAL IMAGING

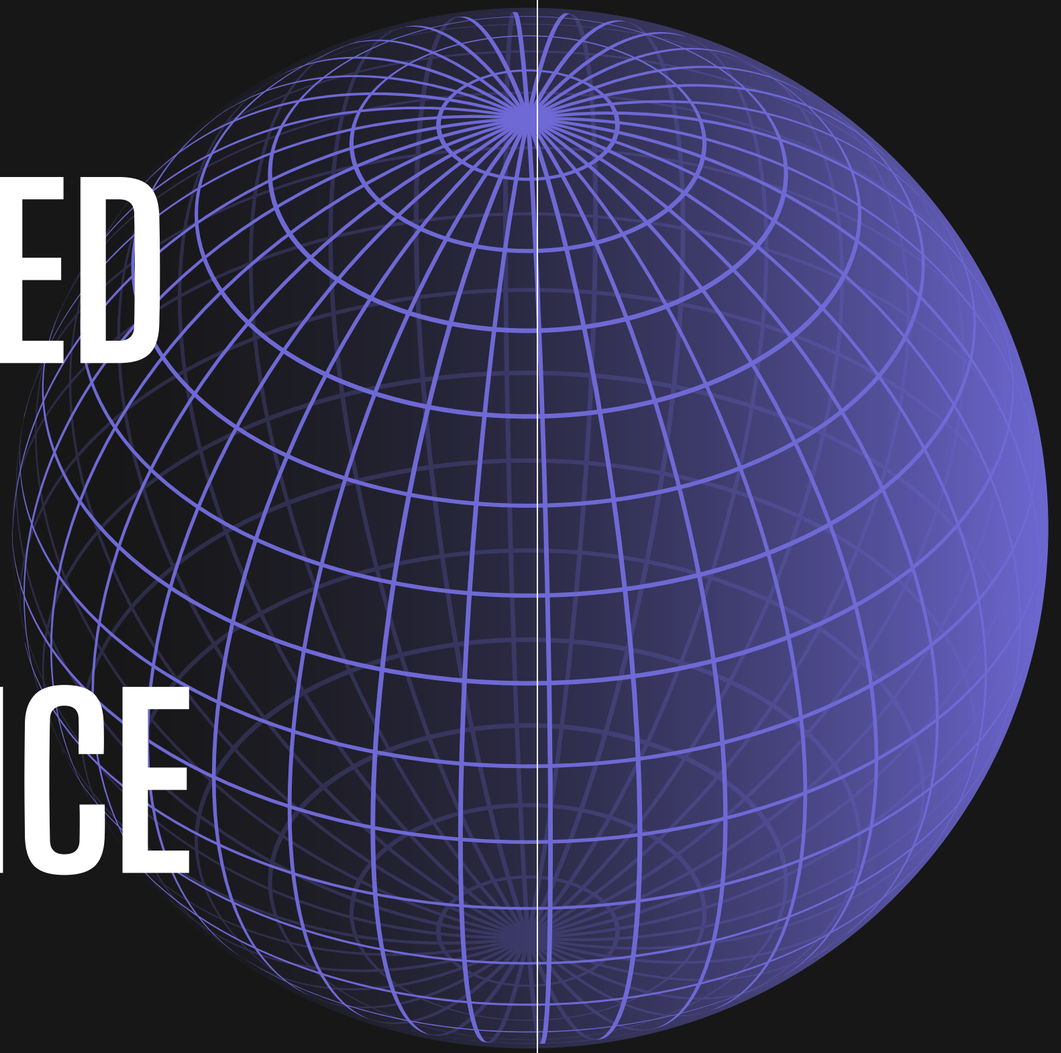
WHY USE DATA SCIENCE IN MEDICAL IMAGING?

MEDICAL DATA, LIKE MRI SCANS, ARE COMPLEX. IT REQUIRES A TRAINED EYE TO INTERPRET, AND EVEN THEN, SUBTLE CHANGES MAY GO UNNOTICED. DATA SCIENCE ALLOWS US TO AUTOMATE THIS PROCESS USING MACHINE LEARNING, WHICH IS FASTER AND OFTEN MORE PRECISE THAN MANUAL METHODS.

HOW DOES MACHINE LEARNING HELP IN DEMENTIA DETECTION?

MACHINE LEARNING MODELS CAN ANALYZE THOUSANDS OF MRI SCANS AND DETECT PATTERNS THAT MAY INDICATE THE EARLY ONSET OF DEMENTIA. THIS EARLY DETECTION CAN BE LIFE-CHANGING BECAUSE IT ALLOWS FOR EARLIER INTERVENTIONS AND BETTER PATIENT OUTCOMES.

APPLIED DATA SCIENCE



IN THE DISCUSED PROJECT:

The model processes an MRI image and classifies it based on the presence and stage of dementia (e.g. non present, mild, moderate, severe).

The predictions made by the model can help doctors make faster, more informed decisions.

BENEFITS OF USING DATA SCIENCE IN HEALTHCARE

EFFICIENCY

Machine learning can analyze large amounts of data quickly.

CONSISTENCY

Unlike humans, models don't get tired or make subjective errors.

BENEFITS OF USING DATA SCIENCE IN HEALTHCARE CONT.

SCALABILITY

Once trained, a model can be deployed to analyze MRI scans worldwide, assisting in diagnosis in regions where specialists are scarce.

DEPLOYMENT: FROM RESEARCH TO REAL-WORLD USE

DEPLOYMENT:

FastAPI for Model Deployment

After training and optimizing the machine learning models, the next step is making them available for real-world use.

In this project, we used FastAPI to create a web service that allows the model to be accessed by doctors or healthcare providers. This service can take in MRI images, process them, and return a prediction about whether dementia is present and, if so, which stage.

THE MODEL IN PRACTICE

User uploads MRI images: A doctor or user can upload an MRI scan to the web application.

Image processing
The system preprocesses the image (resizes, gray-scales, and extracts features).

Prediction
The trained model analyzes the image and predicts whether dementia is present, as well as the stage.

IMPACT ON HEALTHCARE

This kind of deployment makes data science accessible to healthcare professionals without deep technical knowledge. They simply upload the image and receive an automated diagnosis.

THE POWER OF DATA SCIENCE IN HEALTHCARE

KEY TAKEAWAYS:

Data science principles, like preprocessing and machine learning, are critical to analyzing MRI scans and detecting dementia.

Machine learning models, when properly trained and deployed, can assist doctors in making faster and more accurate diagnoses.

The deployment of these models into real-world applications makes them accessible, allowing for broader, scalable healthcare solutions.

CONCLUSION:

Data science has the power to transform healthcare by automating processes and analyzing complex data that would be impossible for humans to process efficiently.

THATS ALL!