



# The Beauty of Machine Learning

A Journey Through IBM—Machine Learning Professional Certificate by Salah

# The Big Picture

---

- One of humanity's most potent tools is mathematics. It enables us to easily conduct our daily activities, such as purchasing and selling. Even when we reached the moon and beyond, these achievements depended entirely on mathematics. Systematic thinking in mathematics allows us to classify objects, predict the future, understand patterns in data, and more. These systematic methods are known as algorithms, which use logic to reach answers. Yet, a significant challenge in some of these algorithms is that they are incredibly lengthy and sometimes impossible to do by hand.
- Therefore, we use computers to make them do calculations for us. Since computers are much faster than us in logic and arithmetic, their capability to apply algorithms to find solutions is remarkable. “It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers” Alan Turing, 1951.

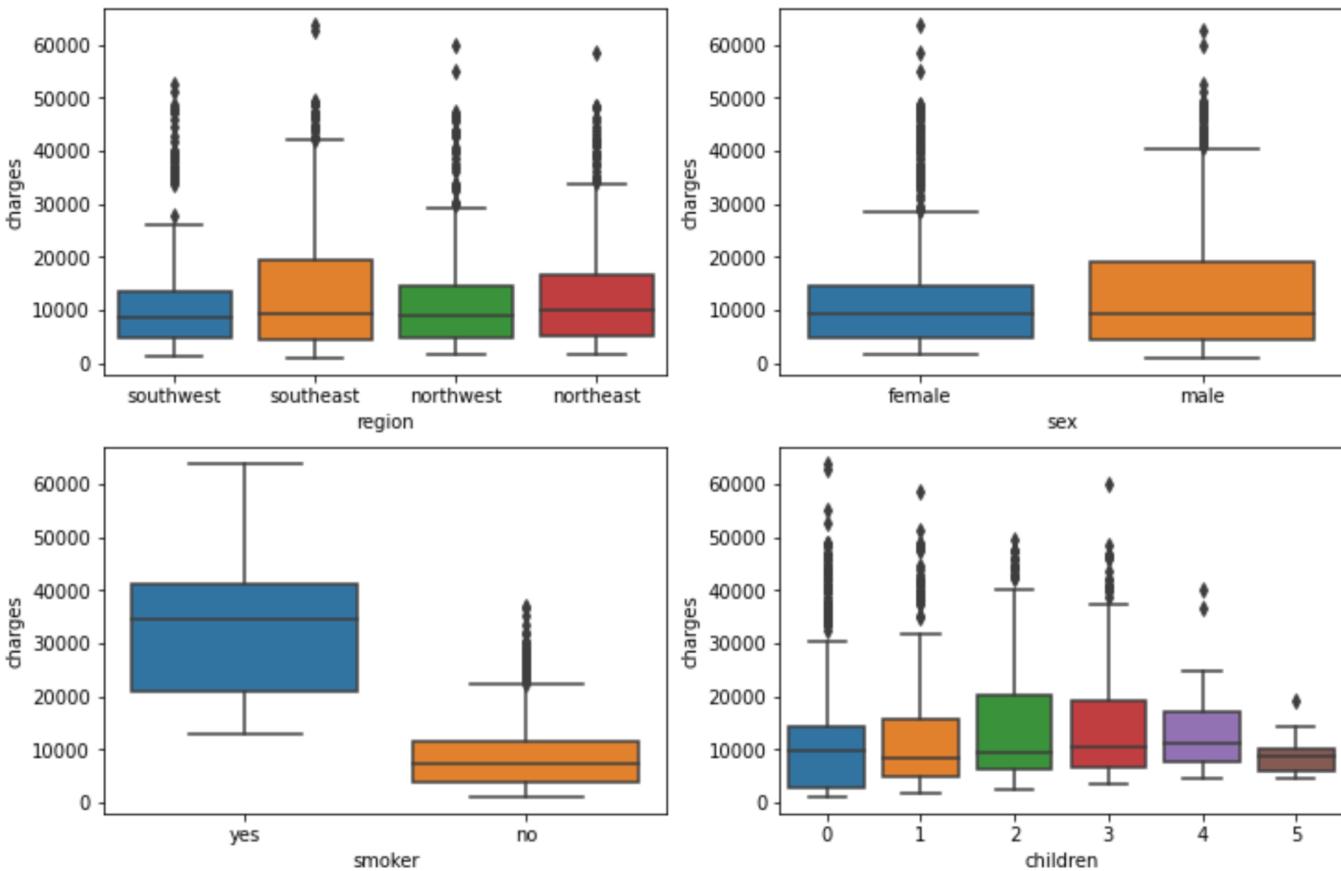
# Machine Learning Algorithms

---

- In short, machine learning algorithms are mathematical approaches to solving problems related to forecasting, classification, clustering, dimensionality reduction, natural language processing, and other applications. Despite the situation, machine learning provides solutions without being explicitly programmed. This beautiful nature of machine learning makes it a great asset in technological advancements.
- In these capstone projects, I used Scikit-Learn framework to build the machine learning models and Keras for the deep learning models. Moreover, these projects cover four sectors: health, green energy, agriculture, and online content. The complexity increases as we move forward from project A to project F.

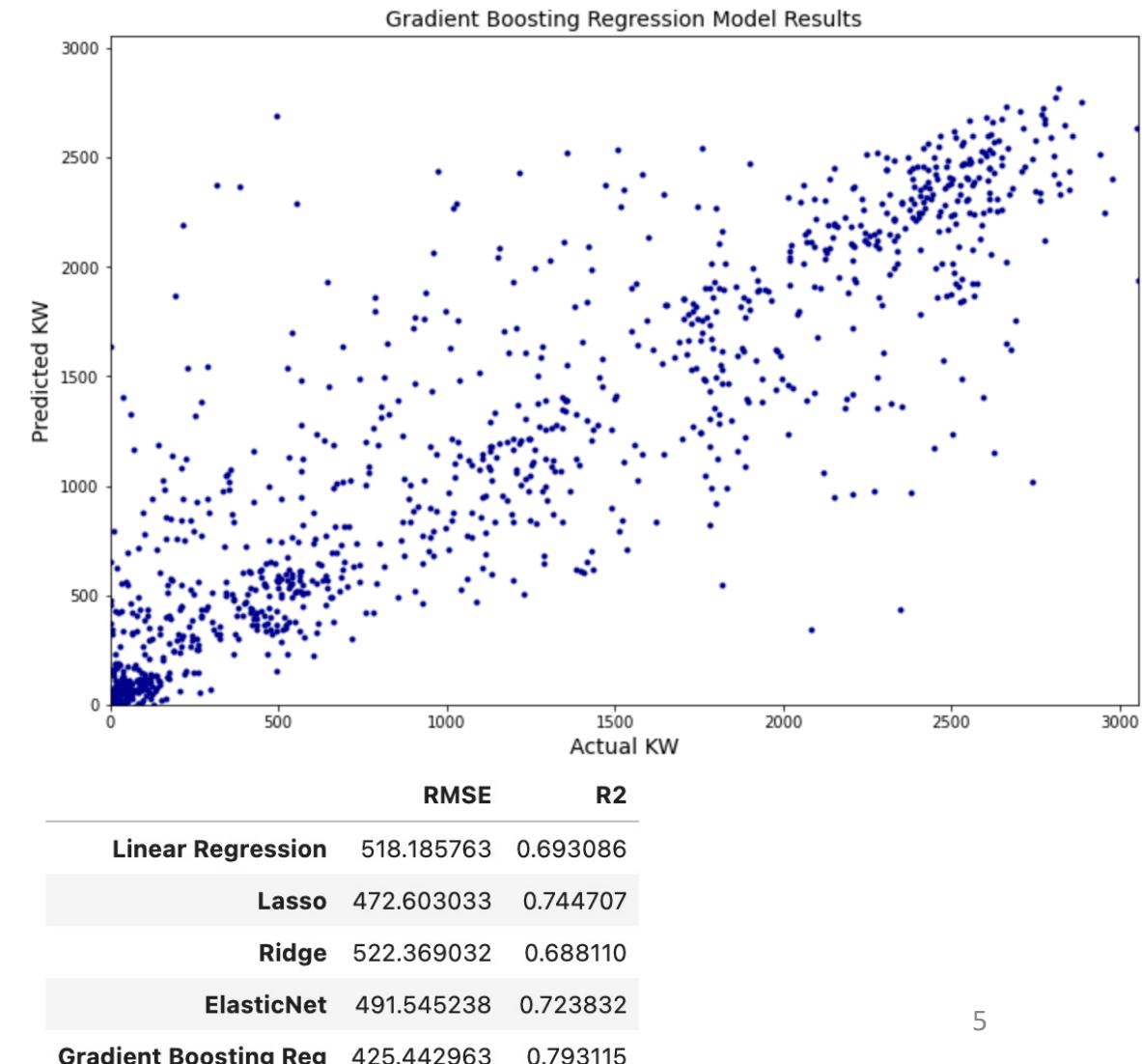
# Project A — Treatment Costs per Person - Exploratory & Predictive Analysis

- Aim: predict the cost of medical treatments based on six features, namely, age, sex, BMI, children, smoking status, and region.
- Procedure: In-depth EDA via pair, bar, box, violin, and regression plots to see the effect of smoking on charges. Hypothesis testing on the relationship between treatment costs and smoking status.
- Findings: The test indicates that a person with a 35K\$ charge or more is likely a smoker with a p-value = 0.023 and a confidence level = 0.977.
- [Project GitHub Notebook](#)



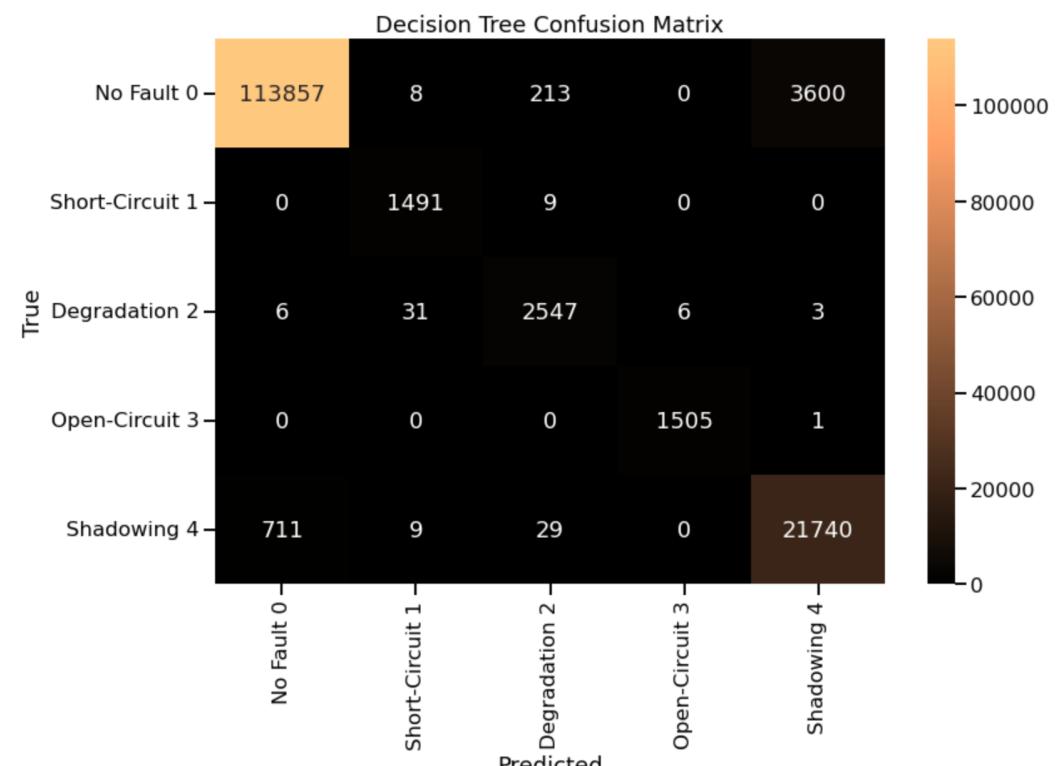
# Project B — Forecasting Photovoltaic Generated Power - Regression Analysis

- Aim: create a regression model that predicts the generated power by PV panels to facilitate energy management in power plants.
- Procedure: Deploy a pipeline encompassing polynomial transformation, standard scaling, and regressor models. Then, apply GridSearchCV, hyperparameters tuning and benchmarking of Regular, Lasso, Ridge, Elastic Net & Gradient Boosting Regressors.
- Findings: The winner is the Gradient Boosting Regressor model with an R2 score of ~ 0.79.
- [Project GitHub Notebook](#)



# Project C — Fault Classification in Photovoltaic Plants - Multi-Class Classification Analysis

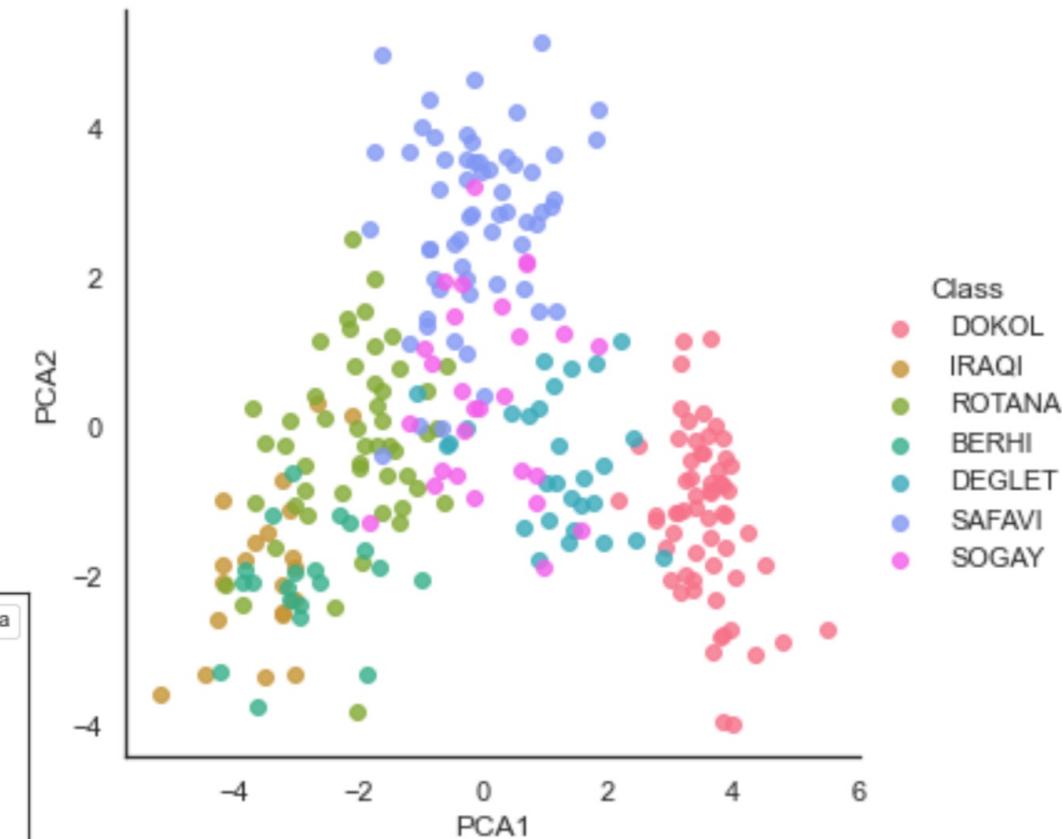
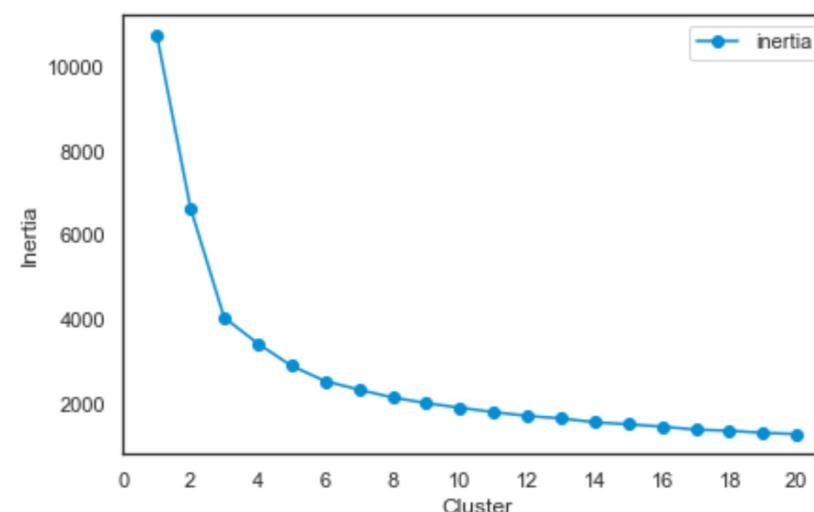
- Aim: Classify the faults that might occur in photovoltaic panels, namely, Short-Circuit, Open-Circuit, Degradation, and Shadowing.
- Procedure: Data stratified split, features scaling, and re-weighting the imbalanced classes. Then, apply a GridSearchCV, hyper-parameters tuning and benchmarking of Logistic Regression, Decision Tree, and Random Forrest.
- Findings: The winner is the Decision Tree algorithm with an accuracy and a weighted F1-score of ~ 97%.
- [Project GitHub Notebook](#)



Metric	Logistic Regression	Decision Tree	Random Forest
Accuracy	0.928221	0.968264	0.949000
Weighted Precision	0.933172	0.971051	0.956368
Weighted Recall	0.928221	0.968264	0.949000
Weighted F1score	0.930049	0.968989	0.950931

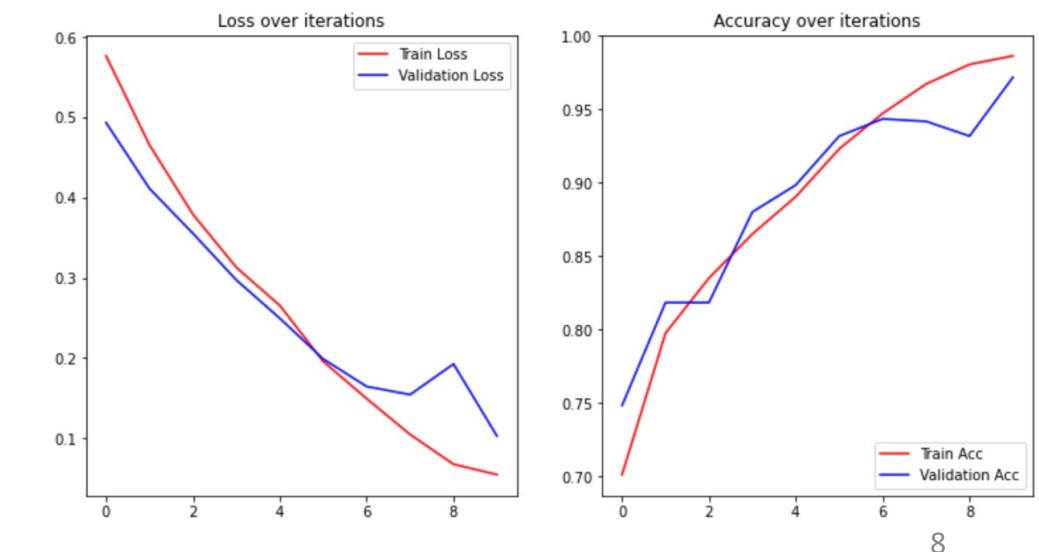
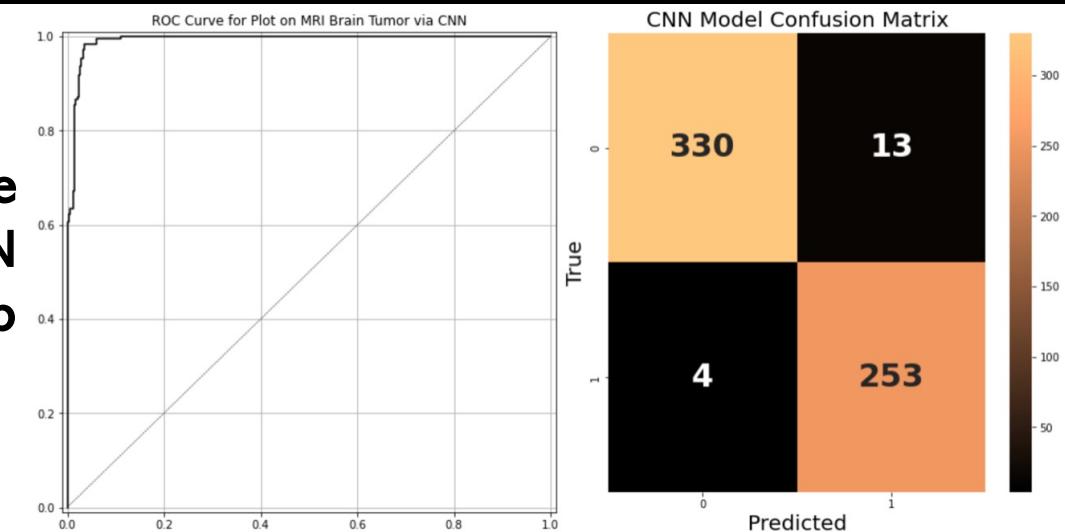
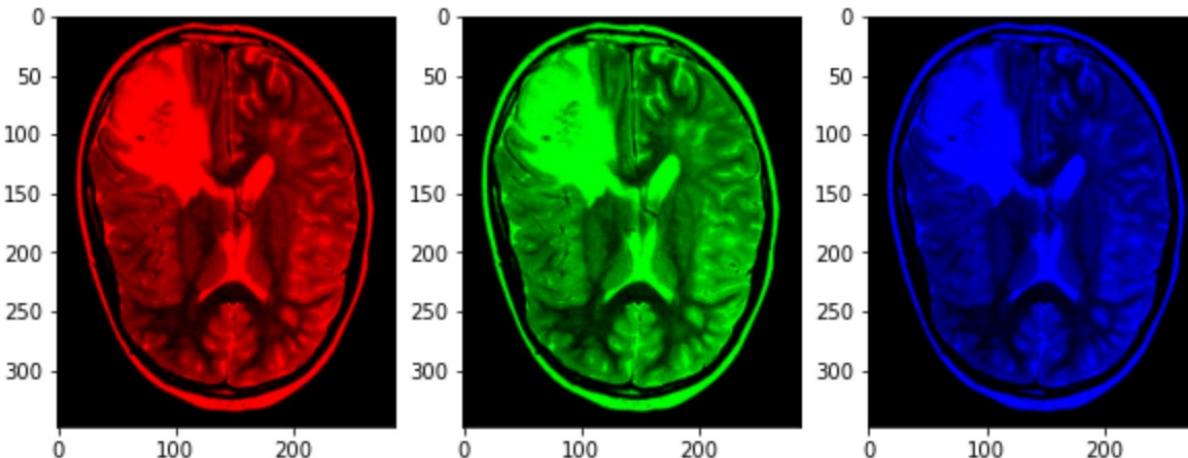
## Project D — Date Fruit Segmentation & Dimensionality Reduction via PCA - Unsupervised Analysis

- Aim: Cluster date fruits based on their physical features.
- Procedure: Check multicollinearity, scale data, and reduce the number of features via PCA. Then, apply a comparative analysis between K-means, Agglomerative, Mean Shift & DBSCAN clustering.
- Findings: The winner is the k-means++ technique. Also, an accuracy of 76% was scored with only two PCAs.
- [Project GitHub Notebook](#)



# Project E — MRI Brain Tumor Classification via CNN - Deep Learning Analysis

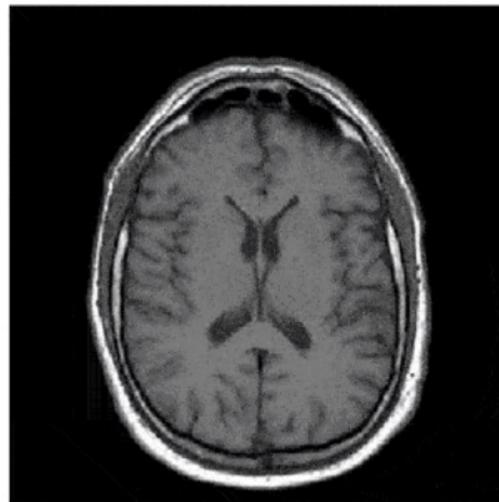
- Aim: Detect whether a patient has a brain tumor or not.
- Procedure: Convert images to a NumPy array and scale them. Build a convolutional network and train the CNN model to classify brain tumors. Then, deploy the deep learning model using Flask app.
- Findings: The CNN model accuracy is 97%.
- [Project GitHub Notebook](#)



# Project E — MRI Brain Tumor Classification via CNN - Dashboard

Brain Tumor Classification Using Deep Learning

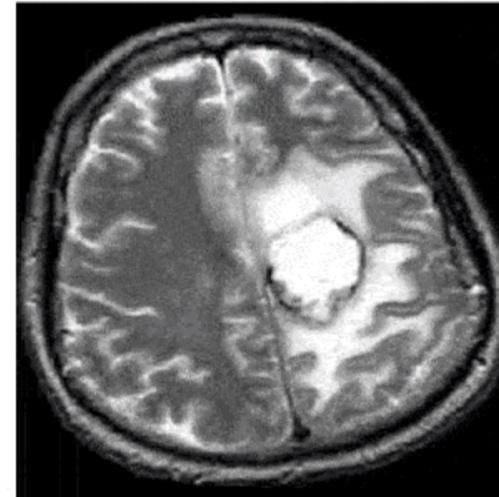
Choose File pred1.jpg



Result: No Brain Tumor

Brain Tumor Classification Using Deep Learning

Choose File pred2.jpg



Result: Yes Brain Tumor

The dashboard is created using Flask framework. First, it prompts the user to select an MRI image. Then, it uses the convolutional neural network model to classify whether there is a brain tumor or not. The two figures demonstrate the outcomes of the model.

# Project F – IBM Capstone Project — Personalized Course Recommendation System for Data Science Learners

- Aim: To build a recommendation system that recommends the most suitable courses for learners on educational platforms.
- Procedure: As listed in the findings, several techniques are used to build the recommendation system.
- Findings: The recommender system is created via eight approaches. Firstly, the content-based approaches.

**Approach 1 - Content-Based Recommender Using User Profile and Course Genres.**

**Approach 2 - Content-Based Recommender Using Course Similarities.**

**Approach 3 - Content-Based Recommender Using PCA Clustering.**

USER	COURSE_ID	SCORE	USER	COURSE_ID	SCORE	user	cluster	Recommended_Unseen_Courses
37465	RP0105EN	27.0	37465	[DS0110EN, excourse67, excourse63, excourse72,...	[0.7329409123199365, 0.7082138557765277, 0.694...	1502801	4	[DS0101EN, PY0101EN, SC0101EN, BD0121EN, DS010...
37465	GPXX06RFEN	12.0	50348	[CB0101EN, DS0110EN, CL0101EN, excourse67, exc...	[0.9233805168766388, 0.7329409123199365, 0.732...	1609720	5	[PY0101EN, DS0103EN, BD0101EN, DS0105EN, ML011...
37465	CC0271EN	15.0	52091	[ML0120ENV3, CB0101EN, excourse24, excourse36,...	[1.0, 0.9233805168766388, 0.7526312050490548, ...	1347188	3	[CO0101EN, CC0101EN, CC0201EN, DS0101EN, BD010...
37465	BD0145EN	24.0	70434	[excouse67, excourse72, excourse74, excourse68]	[0.7082138557765277, 0.7036476305124202, 0.650...	755067	4	[BD0111EN, BD0211EN, BD0115EN, BD0141EN, BD013...
37465	DE0205EN	15.0	85625	[TMP0101EN, excourse72, TA0105EN, BD0151EN, BD...	[0.8894991799933215, 0.7036476305124202, 0.659...	538595	4	[BD0211EN, BD0131EN, DS0101EN, PY0101EN, SC010...
...	...	...	...	...	...	...	...	...
2087663	excouse88	15.0	2061096	[DS0110EN, excourse67, excourse63, excourse72,...	[0.7329409123199365, 0.7082138557765277, 0.694...	481658	4	[BD0211EN, BD0115EN, BD0131EN, DS0101EN, PY010...
2087663	excouse89	15.0	2074313	[excouse36, excouse23, DS0110EN, DV0151EN, e...	[0.7397041774816828, 0.7397041774816828, 0.732...	1341273	3	[CC0101EN, BC0101EN, CO0301EN, CC0103EN, DS010...
2087663	excouse90	15.0	2074462	[CL0101EN, excourse22, excourse62, excourse24]	[0.7327907262791404, 0.6475015976638527, 0.647...	884271	1	[DS0101EN, PY0101EN, DS0103EN, DS0105EN, ML011...
2087663	excouse92	15.0	2082818	[DV0151EN, ML0115EN, ML0122EN]	[0.7235359517703827, 0.7071067811865475, 0.681...	1868084	2	[LB0101Env1, LB0107Env1, BC0101EN]
2087663	excouse93	15.0	2087663	[CB0103EN, excourse24, DS0110EN, CL0101EN, exc...	[0.9233805168766388, 0.7526312050490548, 0.732...	630511	1	[DS0101EN, BD0101EN, DS0103EN, DS0105EN, BC010...

**Approach 1**

**Approach 2**

**Approach 3**

# Project F — IBM Capstone Project — Personalized Course Recommendation System for Data Science Learners

- Findings: The remaining five approaches are collaborative-based. The comparison between them is based on RMSE.

Approach 4 - Collaborative-Filtering Recommender Using K Nearest Neighbor

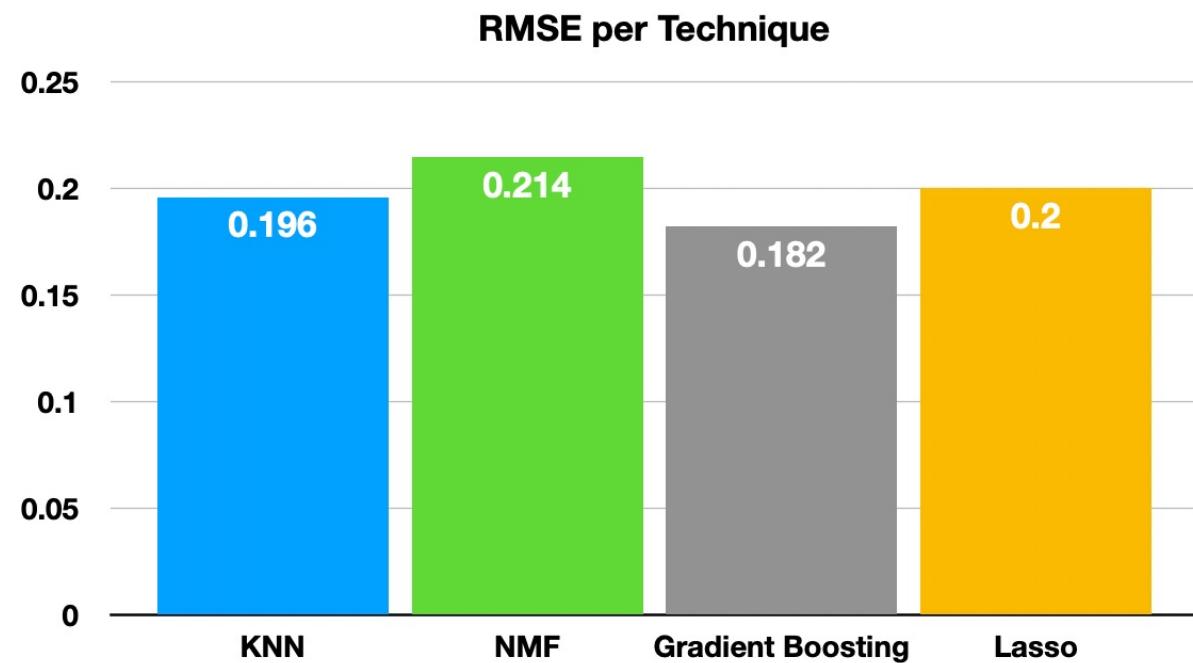
Approach 5 - Collaborative-Filtering Recommender Using Non-negative Matrix Factorization

Approach 6 - Collaborative-Filtering Recommender Using Neural Networks

Approach 7 - Collaborative-Filtering Recommender Using Embedding Features Regression

Approach 8 - Collaborative-Filtering Recommender using Embedding Features

- [Approaches Notebooks](#)



# Project F — IBM Capstone Project — Personalized Course Recommendation System for Data Science Learners Dashboard

**Personalized Learning Recommender**

1. Select recommendation models

Select model:

Course Similarity

2. Tune Hyper-parameters:

Top courses

Course Similarity Threshold %

3. Training:

Train Model

4. Prediction

Recommend New Courses

Made with Streamlit

COURSE_ID	TITLE
0	GPXX0ZGOEN Consuming Restful Services Using The Reactive Jax Rs Client
1	RP0105EN Analyzing Big Data In R Using Apache Spark
2	GPXX0Z2PEN Containerizing Packaging And Running A Spring Boot Application
3	GPXX0FTCEN Learn How To Use Docker Containers For Iterative Development
4	GPXX06RFEN Create Your First Mongodb Database

**Personalized Learning Recommender**

1. Select recommendation models

Select model:

Course Similarity

2. Tune Hyper-parameters:

Top courses

Course Similarity Threshold %

3. Training:

Train Model

4. Prediction

Recommend New Courses

Big Data Modeling

Recommendations generated!

SCORE	TITLE	DESCRIPTION
0 0.5738	Data Analysis Using R 101	data analysis using r 101 teaches you how to perform data analysis using the r language r is a powerful open source language that is ideal for analyzing both structured and unstructured data in this course we will focus on performing analysis of unstructured data files
1 0.5714	Consuming Restful Java Microservices Asynchronously Using Eclipse Microprofile Rest Client	learn how to use microprofile rest client to invoke restful microservices asynchronously over http
		once you ve identified a big data issue to analyze how do you collect store and organize your data using big data solutions in this course you will experience various data genres and management tools appropriate for each you will be able to describe the reasons behind the evolving plethora of new big data platforms from the perspective of big data management systems and analytical tools through guided hands on tutorials you will become familiar with techniques using real time and semi structured data examples systems and tools discussed include asterixdb hp vertica impala neo4j redis sparksql this course provides techniques to extract value from existing untapped data sources and discovering new data sources at the end of this course you will be able to recognize different data elements in your own work and in everyday life problems explain why your team needs to design a big data infrastructure plan and information system design identify the frequent data operations required for various types of data select a data model to suit the characteristics of your data apply techniques to

- The dashboard is created using Streamlit framework. The backend and frontend are created by IBM.



6 Courses

**Exploratory Data Analysis  
for Machine Learning**

**Supervised Machine  
Learning: Regression**

**Supervised Machine  
Learning: Classification**

**Unsupervised Machine  
Learning**

**Deep Learning and  
Reinforcement Learning**

**Machine Learning Capstone**



Aug 26, 2022

## SALAH ALDDEEN YACOUB A. AL KAFRAWI

has successfully completed the online, non-credit Professional Certificate

# IBM Machine Learning

Machine Learning is one of the most in-demand skills for jobs related to modern AI applications, a field in which hiring has grown 74% annually for the last four years (LinkedIn). This Professional Certificate from IBM provides evidence that the above named individual developed skills in Machine Learning: Unsupervised Learning, Supervised Learning, Deep Learning, and Reinforcement Learning; including Time Series Analysis and Survival Analysis. This program consisted of 6 courses providing a solid theoretical understanding and considerable practice of the main algorithms, uses, and best practices related to Machine Learning.

The online specialization named in this certificate may draw on material from courses taught on-campus, but the included courses are not equivalent to on-campus courses. Participation in this online specialization does not constitute enrollment at this university. This certificate does not confer a University grade, course credit or degree, and it does not verify the identity of the learner.

Joseph Santarcangelo  
Senior Data Scientist  
IBM

Mark J Grover  
Digital Content Delivery  
Lead  
IBM Data & AI Learning

Miguel M. Maldonado  
Machine Learning  
Curriculum Developer  
Data and AI Learning

Yan Luo  
Ph.D., Data Scientist  
and Developer

Svitlana Kramar  
Data Science Content  
Developer  
Skills Network

Verify this certificate at:  
<https://coursera.org/verify/profession-al-cert/3P4L22THWW33>