CAPSTONE PROJECT

CORONA VIRUS TWEET SENTIMENT ANALYSIS

Presented By:
1. DIVVELA HEMA HARSHINI-NRI INSTITUTE OF TECHNOLOGY-CSE(AI&ML)

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

- This challenge asks you to build a classification model to predict the sentiment of COVID-19 tweets.

PROPOSED SOLUTION

- **Setup and Initialization**:

  Importing the necessary libraries and setting up the Watson Machine Learning client; configuration of the environment and credentials.

- **Load Data:**

  Loading the COVID-19 tweet dataset into the environment

- **Auto AI Experiment**:

  Definition and run of the Auto AI experiment specifying, for instance, the column to predict, such as sentiment, along with other relevant parameters like the scoring metric or the holdout size. In the AutoAI experiment, several pipelines are explored, including Snap Decision Tree Classifier and Snap Random Forest Classifier, with different optimization steps.

- **Model Selection:**

  Now, among the generated pipelines, choose one with the best performance, usually measured in terms of accuracy or any other evaluation metric. In this case, this will be P8 Snap Random Forest Classifier.

- **Pipeline Optimization:**

  Use hyperparameter tuning and feature engineering on the selected pipeline.

PROPOSED SOLUTION

- **Model Deployment:**

Deploy the optimized model to a cloud environment for online use.

Prepare the deployment by setting a deployment space ID, then configure the deployment metadata.

Create the deployment using the Watson Machine Learning client and get a deployment ID.

- **Test the Deployment**

Test the model in the deployment against new data. Demonstrate that it works according to expectations. Use the deployment endpoint to score new input data and check for the predictions. Maintenance and Cleanup

Delete deployment if needed. List the existing deployments that would help manage and update the deployment space.

SYSTEM APPROACH

Set up the client of IBM Watson Machine Learning, prepare the data, run the Auto AI experiment for generating and selecting of the best pipeline, save and deploy the model, test the deployment, and finally, manage the deployment lifecycle. All of this systematic procedure is in place to efficiently deploy a sentiment analysis model using IBM's cloud-based Auto AI.

ALGORITHM

The experiment has mainly used two different classifiers:

Snap Decision Tree Classifier:

Pipelines: P1, P2, P3, P4

These include various steps of hyper parameter optimization and feature engineering.

Snap Random Forest Classifier:

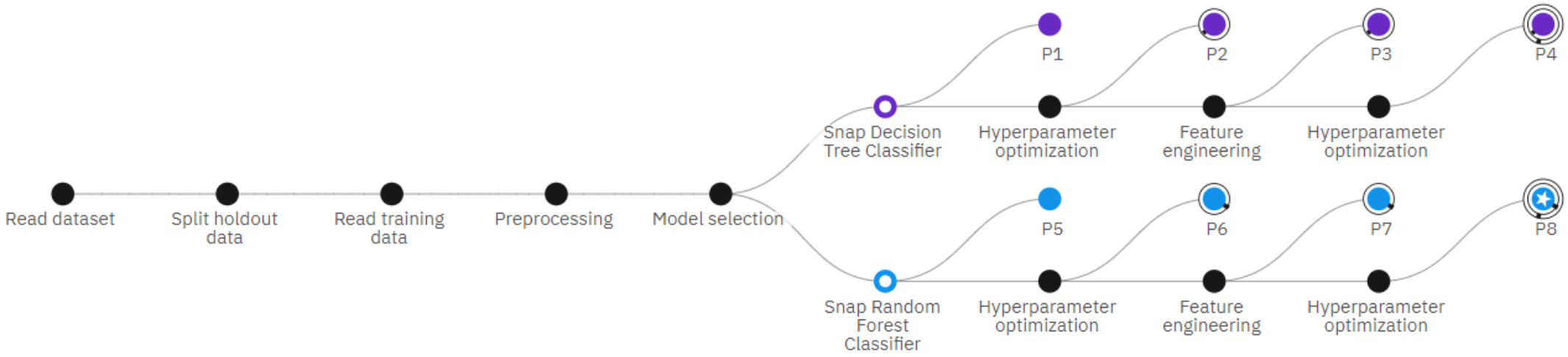
Pipelines: P5, P6, P7, P8

These are also the steps of hyper parameter optimization and feature engineering.

RESULT

Progress map

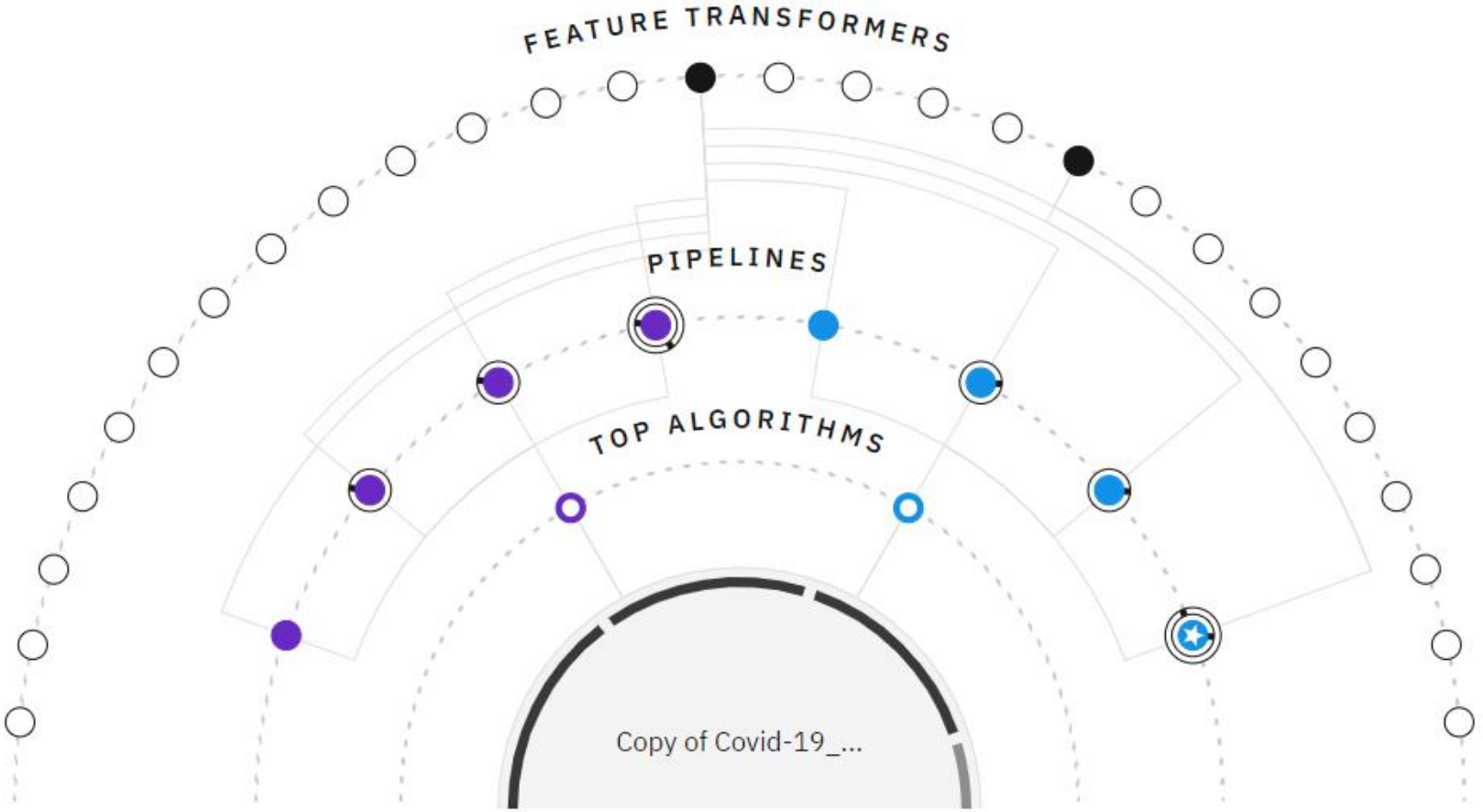
Prediction column: sentiment



RESULT

Relationship map

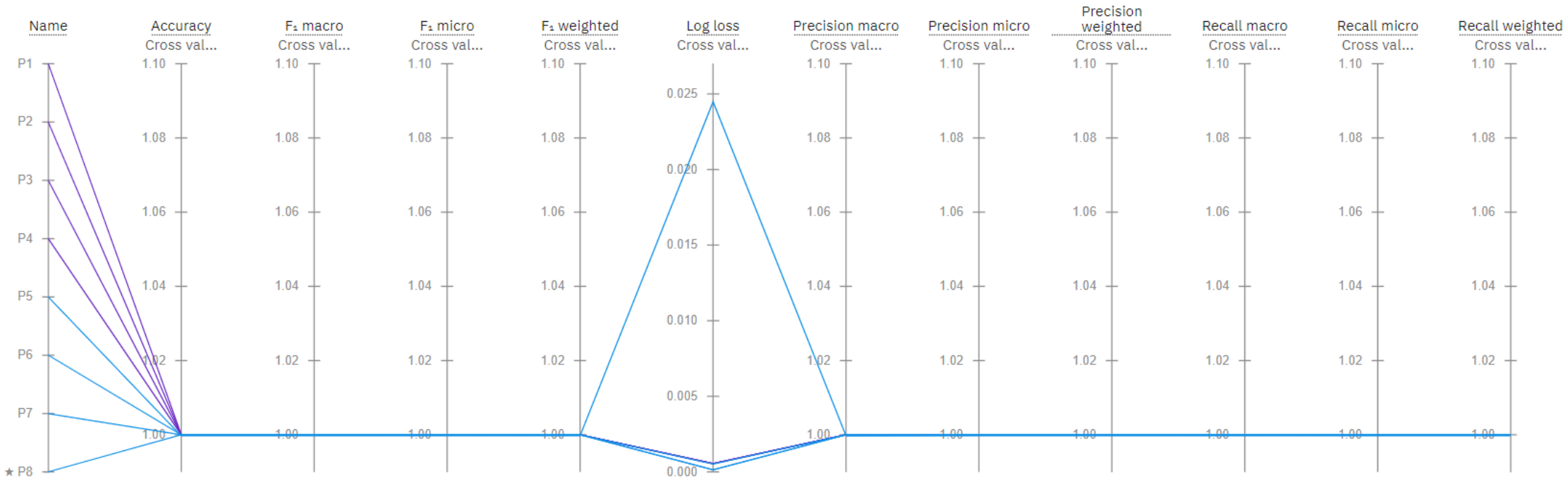
Prediction column: sentiment



RESULT

Metric chart

Prediction column: sentiment



RESULT

	Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	**Pipeline 8**	Snap Random Forest Classifier	**1**	TFE HPO-1 FE HPO-2	00:12:20
	2	**Pipeline 7**	Snap Random Forest Classifier	**1**	TFE HPO-1 FE	00:11:00
	3	**Pipeline 6**	Snap Random Forest Classifier	**1.000**	TFE HPO-1	00:04:46
	4	**Pipeline 4**	Snap Decision Tree Classifier	**1.000**	TFE HPO-1 FE HPO-2	00:03:38
	5	**Pipeline 3**	Snap Decision Tree Classifier	**1.000**	TFE HPO-1 FE	00:02:55
	6	**Pipeline 2**	Snap Decision Tree Classifier	**1.000**	TFE HPO-1	00:01:29
	7	**Pipeline 1**	Snap Decision Tree Classifier	**1.000**	TFE	00:00:40
	8	**Pipeline 5**	Snap Random Forest Classifier	**1.000**	TFE	00:00:47

- The best-performing pipeline (P8) is identified based on its performance during cross-validation.

CONCLUSION

- **Sentiment Analysis**

  Analyze sentiment of COVID-19 tweets, with manual tagging to avoid privacy concerns

- **Model Performance**

  Classify tweet sentiment using Snap Decision Tree Classifier & Snap Random Forest Classifier

- **Future Directions**

  Explore ways to further improve model accuracy and applicability.

FUTURE SCOPE

- This project aims to build a sentiment analysis model for COVID-19 tweets. The tweets have been manually tagged to avoid privacy concerns.

- The goal is to develop a robust system that can accurately classify the sentiment of pandemic-related tweets. This will provide valuable insights into the public's emotional response, informing decision-making and communication strategies.

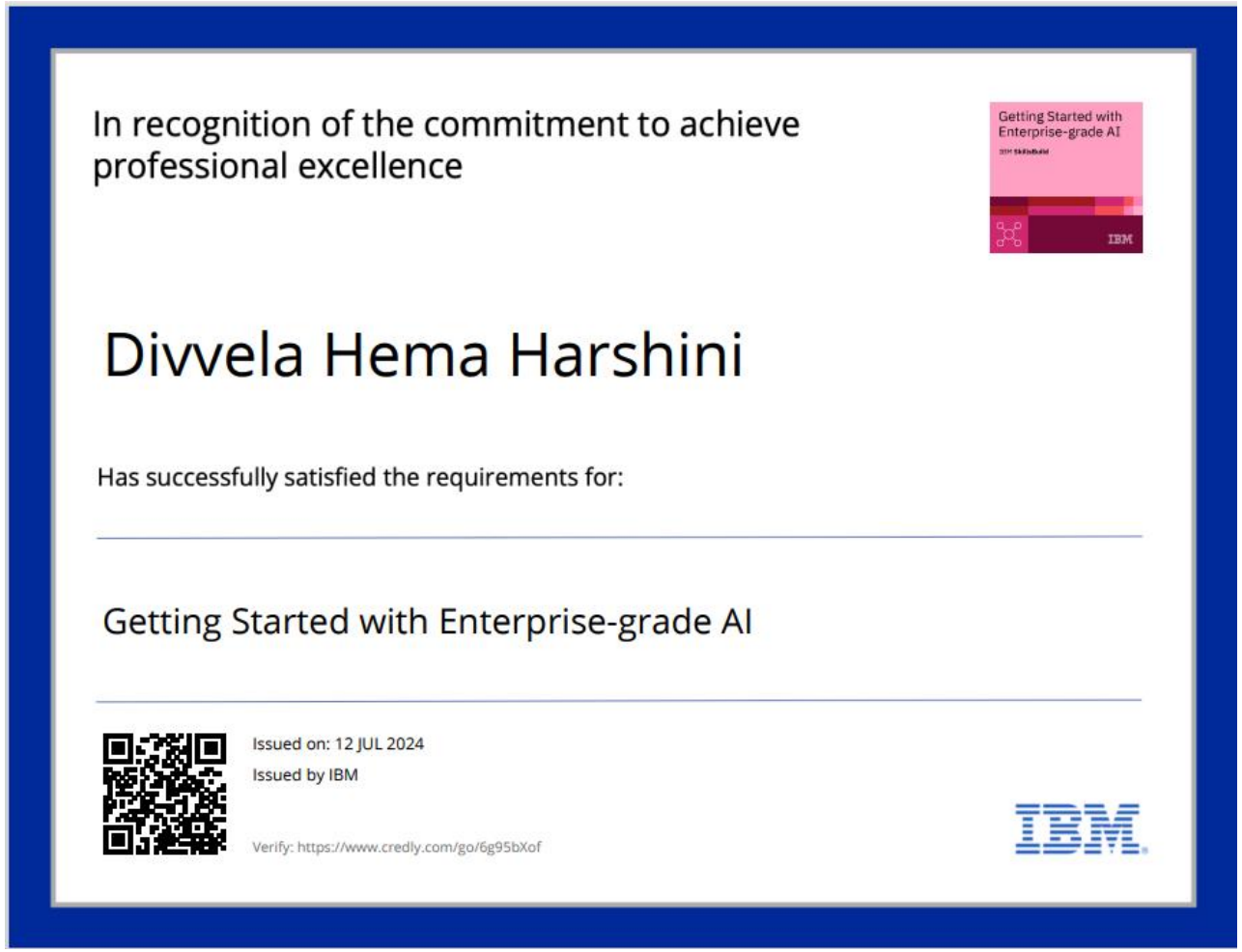
- Beyond the current challenge, there is potential to expand the scope, such as analyzing sentiment trends, identifying key drivers, and exploring regional or demographic differences. The techniques developed can also be adapted to other domains like customer service and brand monitoring.

REFERENCES

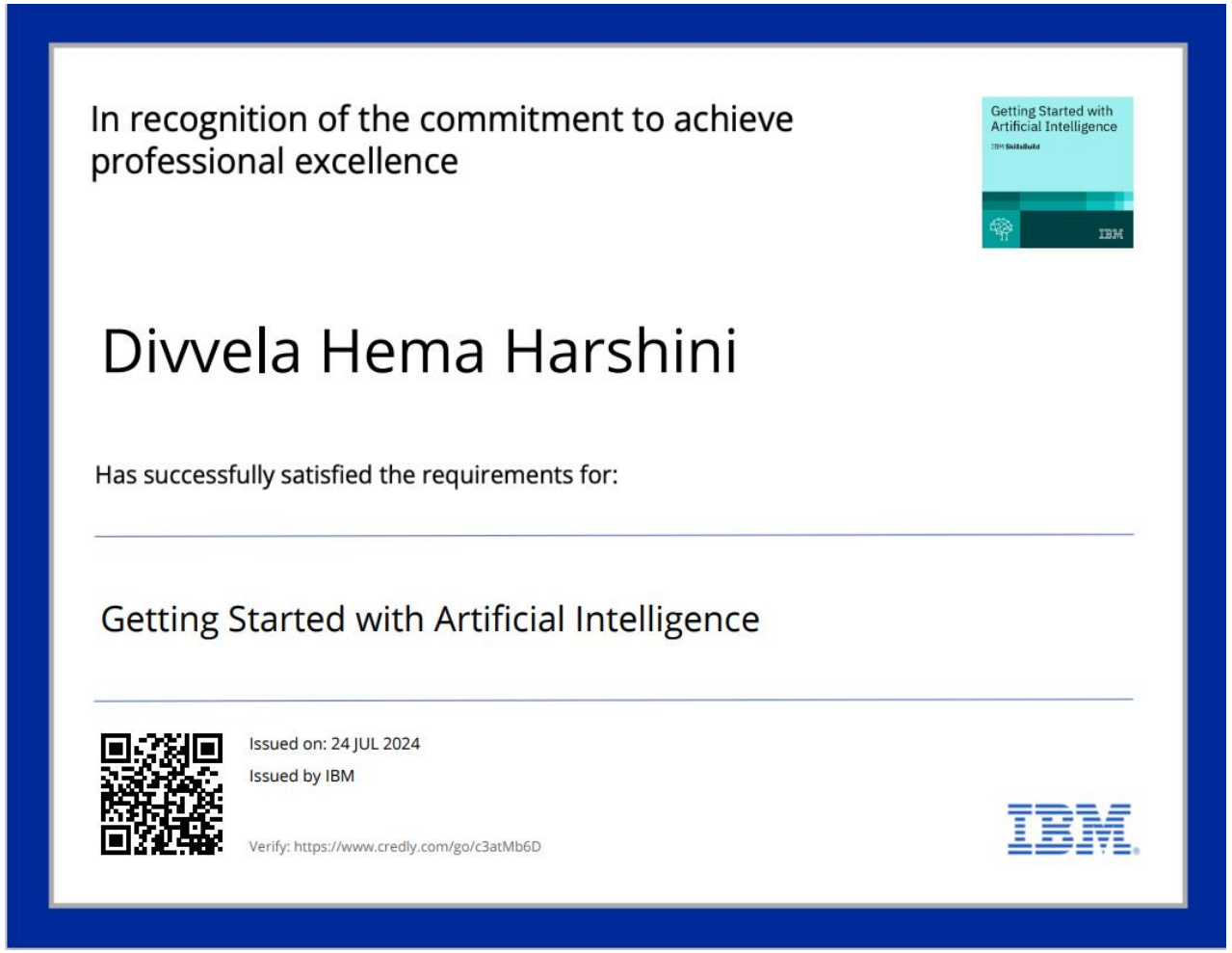
- The data set containing the tweets have been pulled from Twitter and manual tagging has been done then.
- **Google Drive Link:**

https://drive.google.com/drive/folders/1ROKs3PJW-YjxSFpjfYphlXQwVY6mBjWE?usp=sharing

CERTIFICATE1



CETIFICATE 2



THANK YOU

edunet foundation