CAPSTONE PROJECT

PREDICTIVE MAINTENANCE OF INDUSTRIAL MACHINERY

Presented By:
1. Abir Datta-Netaji Subhash Engineering College-CSE

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Unexpected machinery failures lead to costly downtime and inefficient maintenance. This project aims to develop a predictive maintenance model that uses real-time sensor data to classify and anticipate failures—such as tool wear, heat issues, or power loss. By enabling proactive interventions, the model will help reduce downtime and optimize maintenance efforts.

PROPOSED SOLUTION

To reduce unplanned downtime and enable proactive maintenance, we propose developing a machine learning-based predictive maintenance model using the *Kaggle's Machine Predictive Maintenance Classification Dataset* and IBM Watsonx.ai Studio's AutoAI.

The solution would follow a structured pipeline as outlined below:

- <u>**Proposed Workflow**</u>

1. **Data Preparation**
   i. Load and explore the dataset containing sensor readings and failure types.
   ii. Split the data into training and holdout sets.
   iii. Perform preprocessing such as handling missing values, encoding categorical features, and scaling.
2. **Model Development Using IBM AutoAI**
   i. Automatically generate and evaluate multiple models using Snap ML algorithms.
   ii. Focus on:
      - **Snap Decision Tree Classifier**
      - **Snap Random Forest Classifier**
3. **Model Selection and Optimization**
   I. Compare models based on performance metrics (e.g., accuracy).
   II. Select the best-performing model (Snap Random Forest is expected to perform better, achieving up to 99.5% accuracy).

PROPOSED SOLUTION

- **Key Enhancements in the Proposed Model**

  1. **Hyperparameter Optimization:**
     The process of automatically tuning model parameters (such as tree depth, number of estimators, etc.) to find the optimal combination that yields the highest accuracy and avoids overfitting. AutoAI uses intelligent search strategies to achieve this efficiently.
  2. **Feature Engineering:**
     The creation, selection, and transformation of input features to help the model learn more effectively. AutoAI automates this by generating new features, selecting the most informative ones, and removing noise, which enhances predictive performance.

- **Expected Outcome**

  This proposed solution is expected to deliver a high-accuracy classification model capable of predicting machine failures before they occur. By integrating predictive insights into maintenance planning, organizations can reduce downtime, optimize maintenance schedules, and lower operational costs.

SYSTEM APPROACH

The proposed system follows a structured, end-to-end machine learning workflow designed to anticipate machinery failures using sensor data. The methodology includes the following key phases:

- **Data Collection**: Acquire sensor-based machine data from the Kaggle Predictive Maintenance Dataset.
- **Data Preprocessing**: Clean, transform, and prepare the dataset for modeling by handling missing values and encoding features.
- **Model Development**: Use IBM AutoAI to automatically build and compare models using Snap ML algorithms.
- **Model Optimization:** Improve model performance through automated feature engineering and hyperparameter tuning.
- **Model Evaluation:** Assess models using accuracy and other metrics to select the most reliable one.
- **Deployment:** Deploy the final model as an API endpoint on IBM Cloud for real-time usage.

❑ System Requirements:

Watsonx.ai Runtime Service is providing software as well as hardware. For the software part it is creating an instance of Watsonx.ai Runtime and the hardware it is providing is 8 CPU and 32 GB RAM (large) which consumes 20 capacity units.

❑ Libraries used:

Three main services are used to create and deploy the model :

1) Cloud object storage
2) Watsonx.ai runtime
3) Watsonx.ai studio

SYSTEM APPROACH

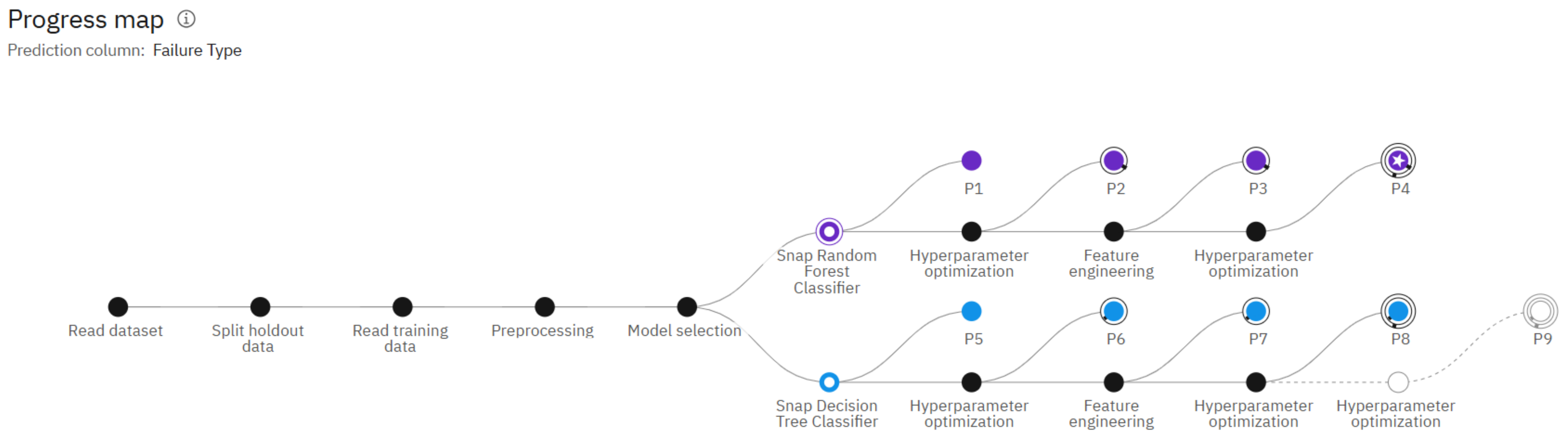


Fig1: Systematic Approach of building the model

- Cloud Object Storage-cc
- watsonx.ai Runtime-db
- watsonx.ai Studio-t1

Services used to develop the model

edunet foundation

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**

Snap Random Forest Classifier is used to develeop the model.
This class implements a random forest classifier using the IBM Snap ML library. It can be used for binary classification problems. It handles both dense and sparse matrix inputs. Use csr, csc or ndarray matrix format for training and csr or ndarray format for prediction.
The **Snap Random Forest Classifier** was chosen for model development due to its superior accuracy and support for advanced features like **automated hyperparameter optimization** and **feature engineering**, which enhance model performance and generalization. Compared to the Snap Decision Tree, it offers better handling of high-dimensional sensor data and complex failure patterns. Its selection is justified as it aligns well with the problem's need for accurate, real-time classification of multiple failure types from diverse operational signals.

- **Data Input:**

Various features are taken as input by the model to predict the fault type such as – UDI , Product ID, Type ,Air Temperature[K] ,Process Temperature[K] , Rotational Speed[rpm] , Torque[Nm] , Tool wear[min] , Target.

ALGORITHM & DEPLOYMENT

- **Training Process:**

The Snap Random Forest Classifier is trained on historical sensor data to learn patterns linked to machine failures. IBM AutoAI enhances this process through **automated feature engineering**, **hyperparameter tuning**, and **cross-validation**, ensuring optimal model performance, reduced overfitting, and reliable predictions on unseen data.
90% of the labeled data which is given as input is used as training dataset and rest is used for test datset.
The model is trained with labeled dataset , where it has been given that for - air temp:298.6 , process temp:309.2 ,rotational speed:1311,torque:46.6,tool wear:44,target:0 , there is no power failure in the machine. There are 10000 instances of data are given , thus the model is trained so that if any random value is given , it can predict the possible outcome with reasonable accuracy.

- **Prediction Process:**

The trained Snap Random Forest Classifier predicts **power failures** by analyzing real-time sensor inputs such as air temp, process temp, rotational speed, torque, tool wear etc. During prediction, this live data is passed to the deployed model via an API, which evaluates it against patterns learned from historical data. The model then classifies whether a power failure is likely to occur, enabling timely maintenance actions.

❑ Deployment:

The project is deployed on the IBM cloud.

1. **Private endpoint:**

https://private.eu-gb.ml.cloud.ibm.com/ml/v4/deployments/4e25bae2-e18b-4714-b7b4-7f2a5d44f8d7/predictions?version=2021-05-01

2. **Public endpoint:**

https://eu-gb.ml.cloud.ibm.com/ml/v4/deployments/4e25bae2-e18b-4714-b7b4-7f2a5d44f8d7/predictions?version=2021-05-01

RESULT

This machine learning model which is developed using Snap random classifier gives an accuracy of **99.5%.**

Pipeline leaderboard

	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	**Pipeline 4**	Snap Random Forest Classifier		0.995	HPO-1 FE HPO-2	00:00:40
	2	**Pipeline 3**	Snap Random Forest Classifier		0.995	HPO-1 FE	00:00:30
	3	**Pipeline 8**	Snap Decision Tree Classifier		0.994	HPO-1 FE HPO-2	00:00:33
	4	**Pipeline 2**	Snap Random Forest Classifier		0.994	HPO-1	00:00:09

Fig2: Selection of the classifier based on highest accuracy

Observed	Predicted						
	Heat Dissipation Failure	No Failure	Overstrain Failure	Power Failure	Random Failures	Tool Wear Failure	Percent correct
Heat Dissipation Failure	10	0	1	0	0	0	90.9%
No Failure	0	965	0	0	0	0	100.0%
Overstrain Failure	0	0	8	0	0	0	100.0%
Power Failure	0	0	0	10	0	0	100.0%
Random Failures	0	2	0	0	0	0	0.0%
Tool Wear Failure	0	0	0	0	0	4	100.0%
Percent correct	100.0%	99.8%	88.9%	100.0%	0.0%	100.0%	99.7%

Less correct — More correct

Fig3: Confusion Matrix

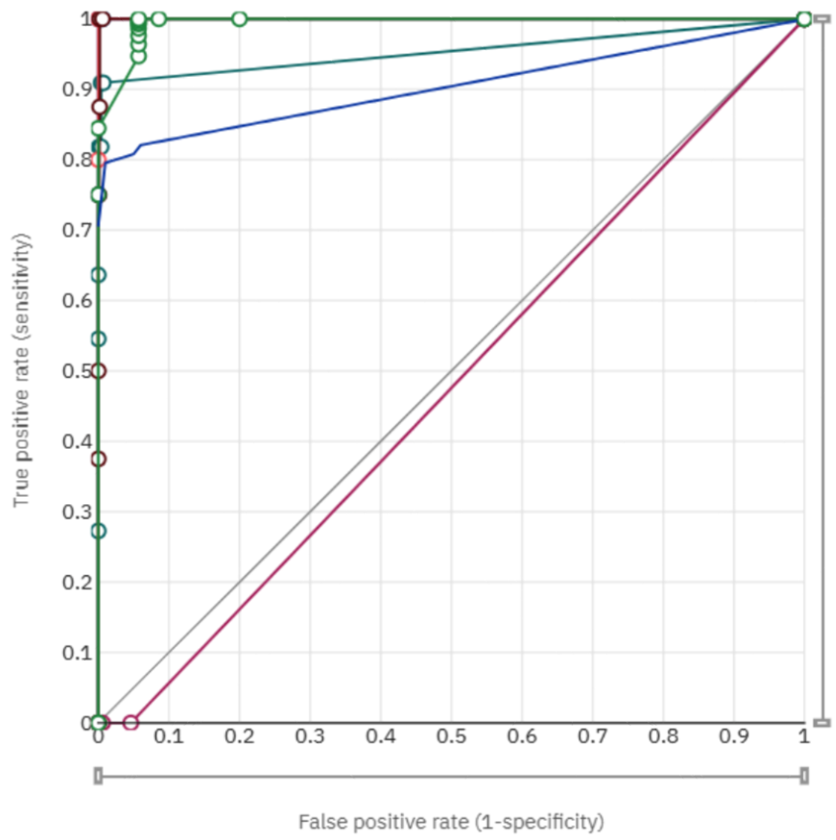
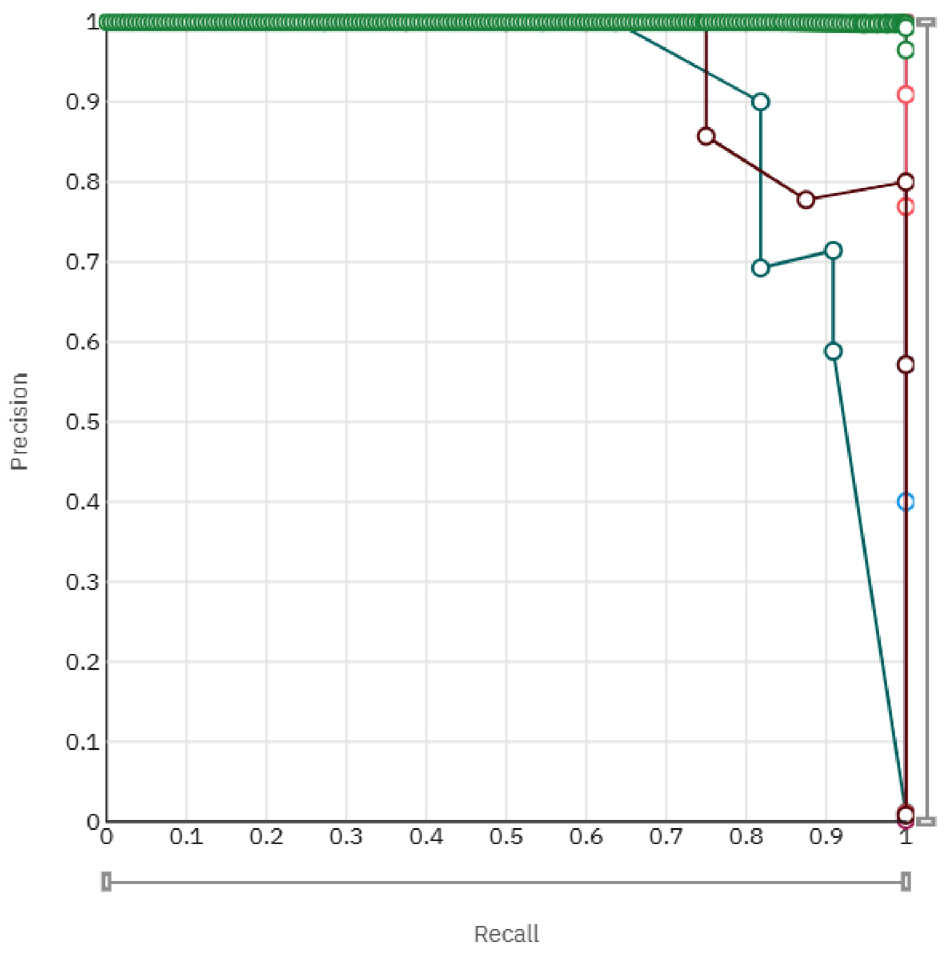


Fig4: ROC Curve



<u>Fig5: Precision Recall Curve</u>

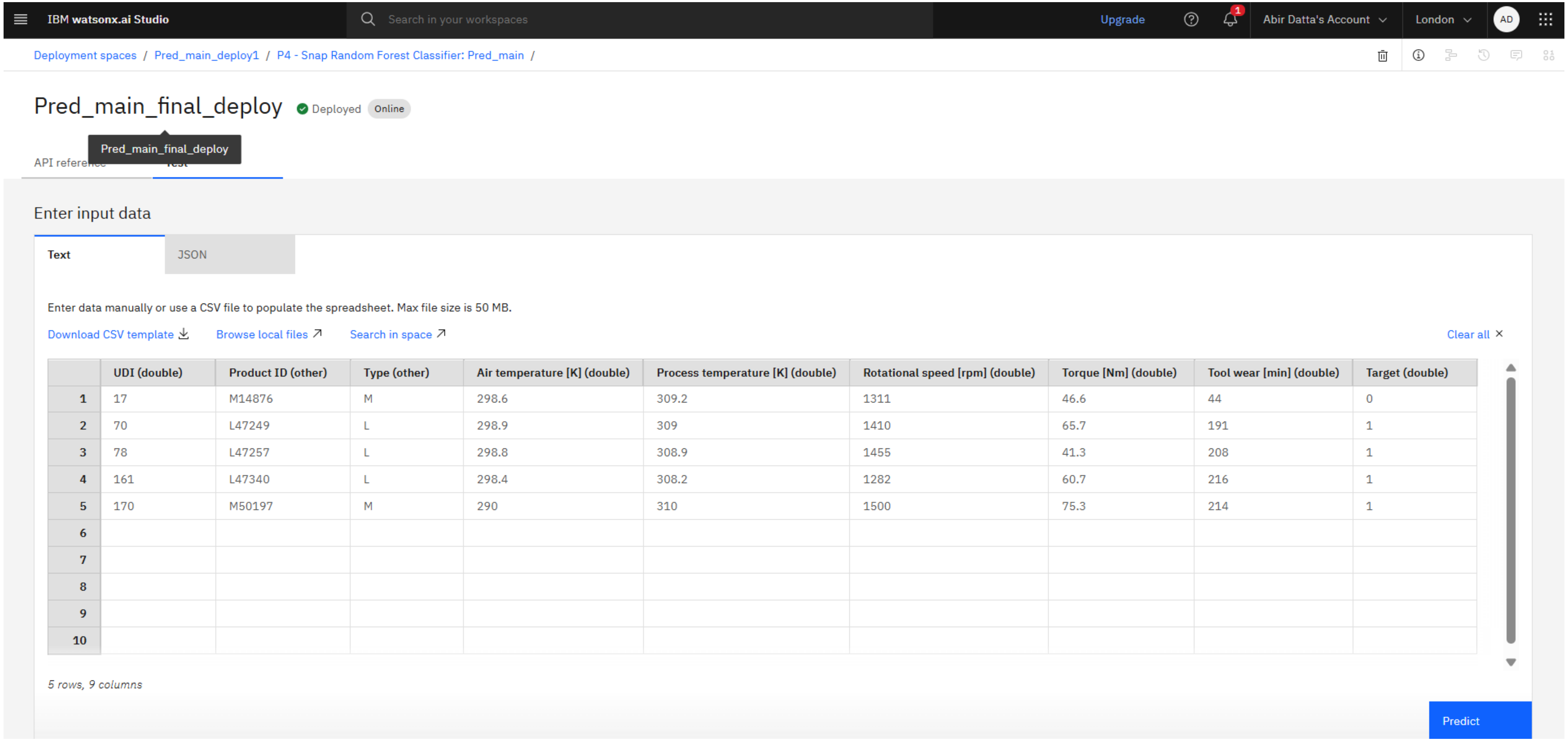


Fig6: Input data in the final model for testing

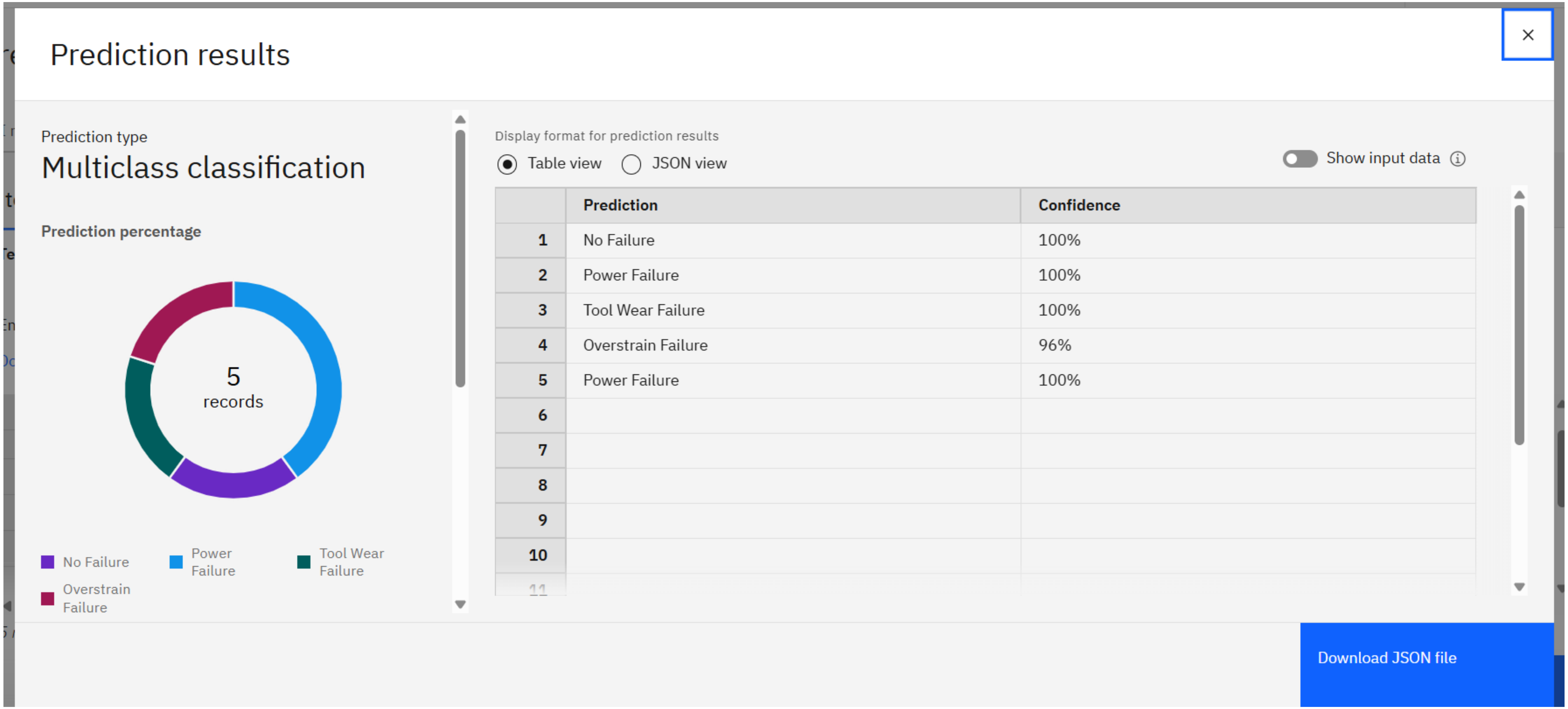


Fig7: Prediction results based on the input data

CONCLUSION

Using the Snap Random Forest Classifier, the proposed predictive maintenance method achieved 99.5% accuracy in predicting machine failure categories such as heat overstrain, tool wear, and power failure. The model accurately predicts failures before they occur by using real-time sensor values and auto-improvements like feature engineering and hyperparameter adjustment.

During implementation, IBM AutoAI's automation and optimization capabilities were used to address issues like class imbalance and complex data patterns.

In summary, the solution shows that precise, real-time failure prediction is not only possible but also very successful in facilitating proactive maintenance, which dramatically lowers industrial machinery operating costs and downtime.

FUTURE SCOPE

To further improve the system's effectiveness and scalability, several enhancements can be considered:

- **Incorporating Additional Data Sources:**
  Integrate real-time inputs such as environmental data, maintenance logs, and operator feedback to enrich model accuracy.
- **Algorithm Optimization:**
  Explore advanced machine learning techniques like ensemble stacking or deep learning to improve prediction precision and handle more complex failure patterns.
- **Geographic Expansion:**
  Scale the system to support predictive maintenance across multiple cities or industrial regions by deploying cloud-based instances and localized models.
- **Edge Computing Integration:**
  Deploy models at the edge (on or near machines) for faster, low-latency predictions without relying on constant cloud connectivity.

These enhancements can significantly boost the system's performance, responsiveness, and adaptability in large-scale industrial applications.

REFERENCES

1. https://www.sap.com/india/products/scm/apm/what-is-predictive-maintenance.html

2. https://ieeexplore.ieee.org/document/9509696/

3. https://www.researchgate.net/publication/337762244_Predictive_Maintenance_and_Monitoring_of_Industrial_Machine_using_Machine_Learning

4. https://www.3ds.com/industries/industrial-equipment/predictive-maintenance-manufacturing-resilience

5. https://ijisae.org/index.php/IJISAE/article/view/4689

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Abir Datta

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 15, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/979c49a7-a929-41df-91bf-9f17afb34e31

IBM

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Abir Datta

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/5506cbec-1ace-4efe-85f5-1752058c1eb6

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Abir Datta

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT) **Learning hours:** 20 mins

THANK YOU