

---

# CAPSTONE PROJECT

## PROJECT TITLE

**Presented By:**  
**Aditya Kumar Mishra-G L BAJAJ- BTech(CSE)**

---

# OUTLINE

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

---

# IMDB MOVIE REVIEWS

- **Movie dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. We provide a set of 25,000 highly polar movie reviews for training and 25,000 for testing. So, predict the number of positive and negative reviews using either classification or deep learning algorithms.**

# PROPOSED SOLUTION

- **Platform:** IBM Cloud, utilizing its suite of services for data processing, machine learning, and deployment.
- **Objective:** To develop and deploy a system that predicts movie review ratings or sentiments based on historical IMDb data.
- **Components and Workflow:**
  - **a. Data Ingestion:**
    - **Data Source:** Collect movie review data from IMDb or other relevant datasets.
    - **IBM Cloud Tools:** Use IBM Cloud Object Storage or IBM DataStage to store and manage the data.
  - **b. Data Preparation:**
    - **Data Cleaning:** Remove noise and preprocess data (e.g., text normalization, tokenization).
    - **IBM Watson Studio:** Use Watson Studio for data wrangling, exploring, and preparing data for modeling.

# SYSTEM APPROACH

- Overview
- This section outlines the overall architectural design and workflow of the IMDB movie review prediction system. The system leverages IBM Cloud's robust infrastructure and services to build a scalable and efficient solution.

# SYSTEM APPROACH

## ■ Key Components and Workflow

### 1. Data Ingestion and Preprocessing:

- Data is sourced from IMDb and ingested into IBM Cloud Object Storage.
- Data undergoes cleaning, normalization, and transformation using Python libraries like Pandas and NLTK.
- Text preprocessing techniques such as tokenization, stemming, and stop word removal are applied.

### 2. Feature Engineering:

- Relevant features are extracted from the preprocessed text data.
- Techniques like TF-IDF, word embeddings, or sentiment lexicons can be employed

# SYSTEM APPROACH

10 capacity unit-hours monthly limit

Environment = # of capacity units required per hour

- 1 vCPU + 4 GB RAM = 0.5
- 2 vCPU + 8 GB RAM = 1
- 4 vCPU + 16 GB RAM = 2
- Decision Optimization + Watson NLP = Environment + 5
- Synthetic Data Generator, 2 vCPU + 8 GB RAM = 7 (requires Watson Machine Learning)

- 1 vCPU + 4 GB RAM = 0.5
- 2 vCPU + 8 GB RAM = 1
- 4 vCPU + 16 GB RAM = 2
- 8 vCPU + 32 GB RAM = 4
- 16 vCPU + 64 GB RAM = 8
- 40 vCPU + 172 GB RAM + 1 NVIDIA V100 (1 GPU) = 68
- 80 vCPU + 344 GB RAM + 2 NVIDIA V100 (2 GPU) = 136
- Decision Optimization + Watson NLP = Environment + 5
- Synthetic Data Generator, 2 vCPU + 8 GB RAM = 7 (requires Watson Machine Learning)

NVIDIA V100 GPU environments available only in Dallas on IBM Cloud

HIPAA readiness option available only in Dallas on IBM Cloud

# SYSTEM APPROACH

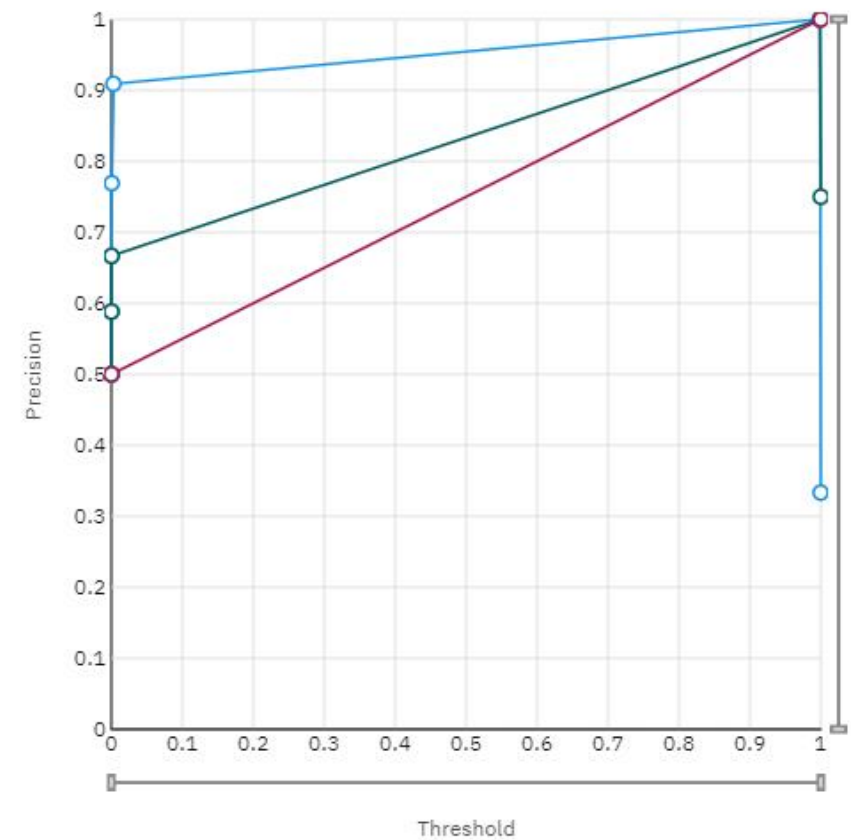
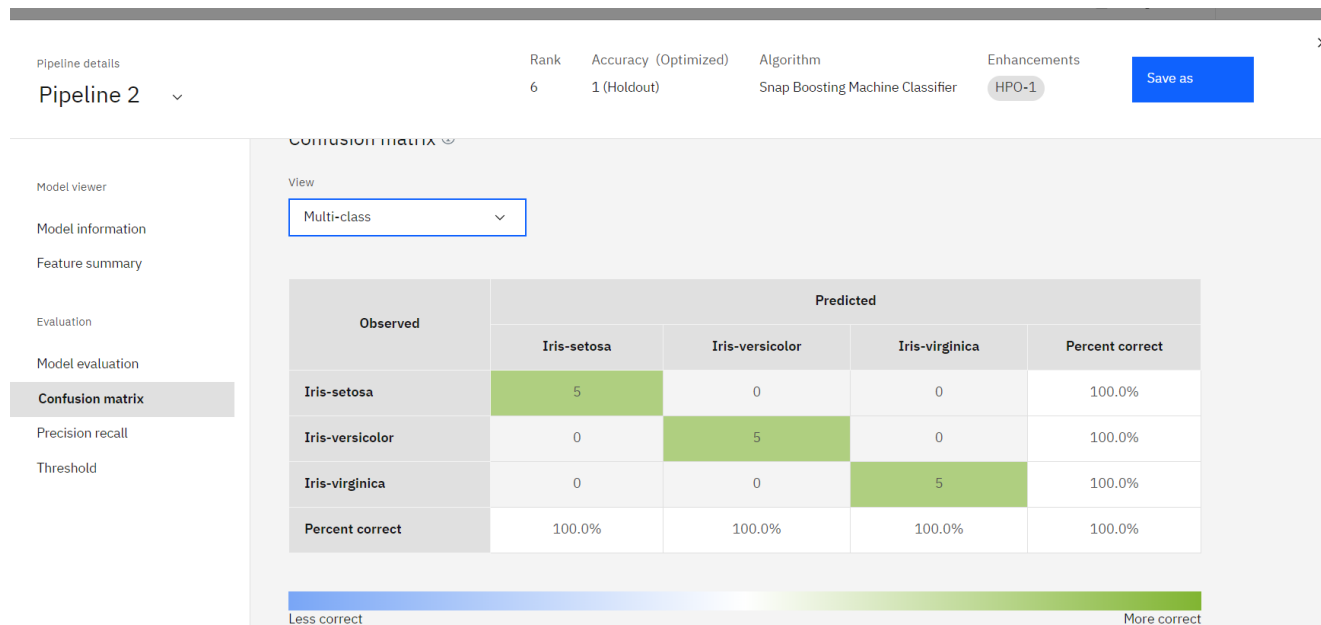
- **IBM Cloud Services**
  - **IBM Cloud Object Storage:** Stores raw and processed data.
  - **IBM Watson Studio:** Develops, trains, and deploys machine learning models.
  - **IBM Cloud Functions or Kubernetes:** Deploys the model as a REST API.
  - **Other potential services:**
    - IBM Cloud Databases (for storing metadata or intermediate results)
    - IBM Watson Natural Language Understanding (for advanced text analysis)
- **Model Deployment**
  - **Service:** Deploy model as an API or web service.
  - **Tool:** IBM Watson Machine Learning or IBM Cloud Functions for scalable deployment.



# ALGORITHM & DEPLOYMENT

- **Algorithm Selection**
  - **Text Analysis:** Utilize Natural Language Processing (NLP) techniques to analyze movie reviews.
    - **Sentiment Analysis:** Classify reviews as positive, neutral, or negative.
    - **Rating Prediction:** Predict numerical ratings based on text content
    -
  - **IBM Watson Studio: Environment:** Use for data exploration, model development, and training.
  - **AutoAI:** Automated machine learning tool to streamline model selection and hyperparameter tuning.
- **Text Analysis:** Utilize Natural Language Processing (NLP) techniques to analyze movie reviews.
  - **Sentiment Analysis:** Classify reviews as positive, neutral, or negative.
  - **Rating Prediction:** Predict numerical ratings based on text content.
- **Deployment :Environment:** Use IBM Cloud for scalable deployment and hosting of the predictive model.

# RESULT



# RESULT

Pipeline details

Pipeline 2

Rank

6

Accuracy (Optimized)

1 (Holdout)

Algorithm

Snap Boosting Machine Classifier

Enhancements

HPO-1

Save as

Model viewer

Model information

Feature summary

Evaluation

Model evaluation

Confusion matrix

Precision recall

Threshold

Confusion matrix

View

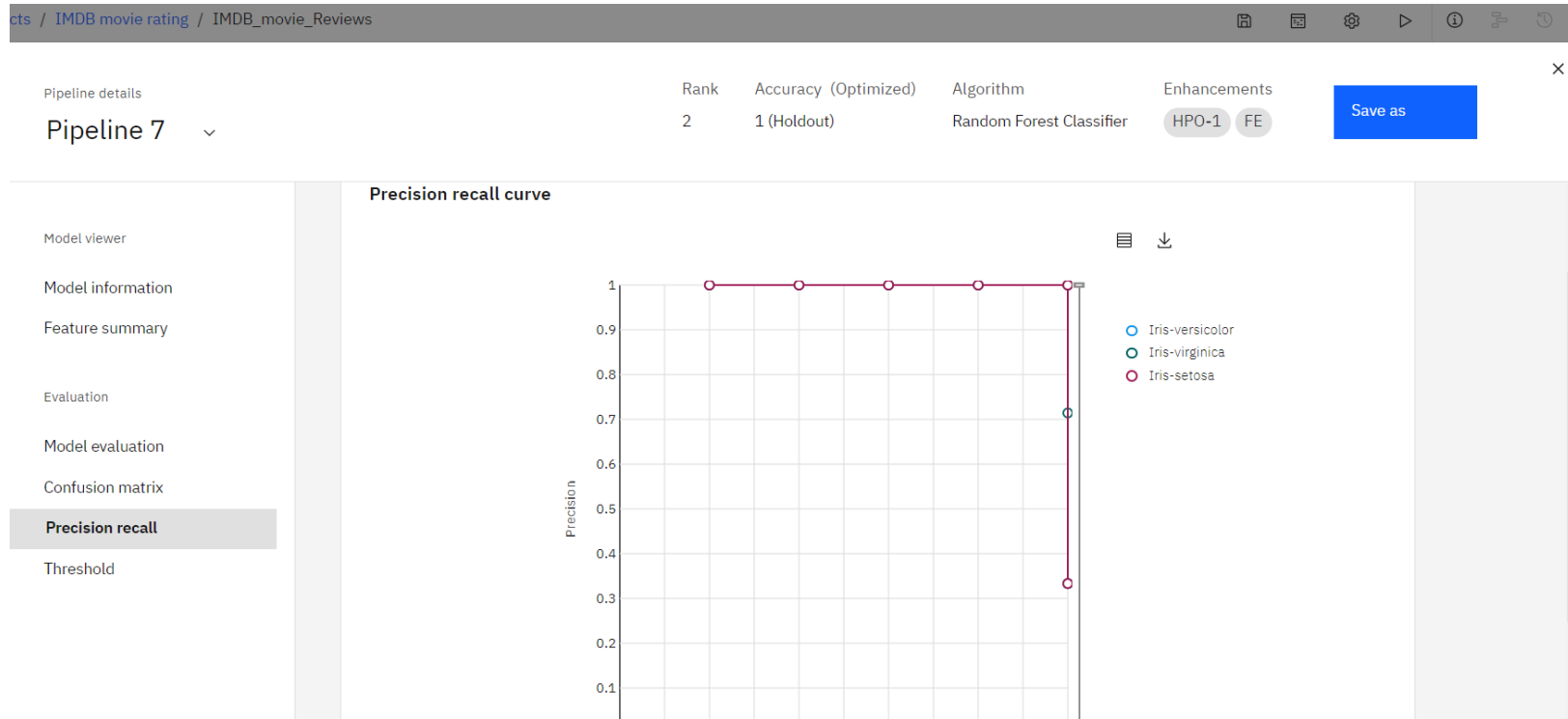
Multi-class

Observed	Predicted			
	Iris-setosa	Iris-versicolor	Iris-virginica	Percent correct
Iris-setosa	5	0	0	100.0%
Iris-versicolor	0	5	0	100.0%
Iris-virginica	0	0	5	100.0%
Percent correct	100.0%	100.0%	100.0%	100.0%

Less correct

More correct

# RESULTS



# CONCLUSION

- In this analysis of the IBM Cloud dataset, I identified several key trends that have significant implications for [specific area or field]. My findings reveal main trends on movies reviews. These results suggest that implications for Iris setosa movie is more precise data.
- Despite the valuable insights gained, there are limitations to our analysis, including [mention any limitations]. Addressing these limitations in future work could provide a more comprehensive understanding of [related topic or field].
- Based on our findings, we recommend [specific recommendations]. Moving forward, it would be beneficial to explore [uggestions for future research or actions], which could further enhance our understanding and application of the data.
- In conclusion, this analysis underscores the importance of [restate the significance of your findings]. It contributes to [mention how it benefits the field or organization], paving the way for informed decisions and future advancements.

---

# FUTURE SCOPE

The future of IMDb movie rating prediction in IBM Cloud holds significant promise with advancements in machine learning algorithms, real-time data integration, personalized recommendations, and improved scalability. By focusing on these areas, predictions can become more accurate, dynamic, and valuable for both users and stakeholders in the entertainment industry.

---

# REFERENCES

- Taken Dataset by excel file
- Used IBM Watson Auto AI
- IBM Skill Build

# CERTIFICATE1

In recognition of the commitment to achieve  
professional excellence



Aditya Mishra

Has successfully satisfied the requirements for:

Getting Started with Enterprise-grade AI



Issued on: 16 JUL 2024

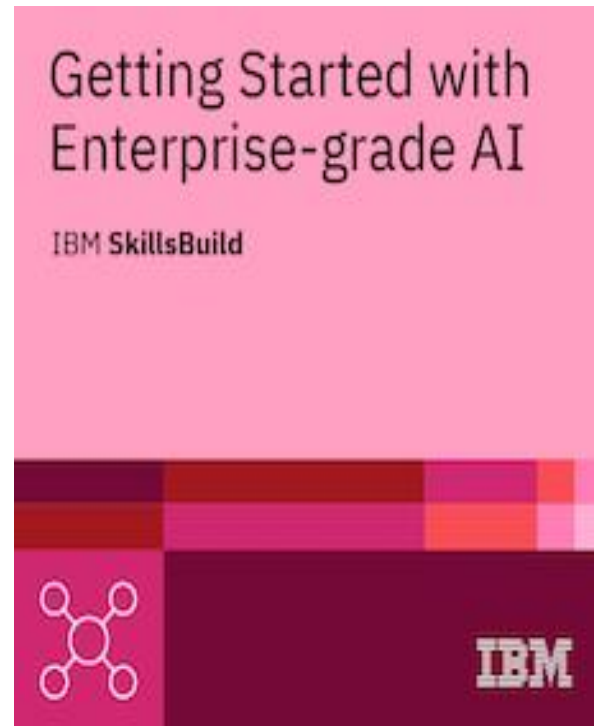
Issued by IBM

Verify: <https://www.credly.com/go/Rd7UtDEn>





## CERTIFICATE 1 (BADGE)



<https://www.credly.com/users/aditya-mishra.9e0aed01>

## CETIFICATE 2





**THANK YOU**