

---

# CAPSTONE PROJECT

## AMAZON SALES DATA PREDICTION

**Presented By:**

- 1. Student Name- DEEPAK BEHERA**
- 2. College Name- Sai Balaji International Institute of Management Sciences**
- 3. Department- Business Analytics**

---

# OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result**
- **Conclusion**
- **Future Scope**
- **References**

# PROBLEM STATEMENT

**Example:** Retail businesses often face significant challenges in managing inventory, planning marketing strategies and forecasting sales due to the dynamic nature of consumer demand and market conditions. Inaccurate sales predictions can lead to the stockouts, excess inventory, missed revenue opportunities and increased operational costs.

- ❑ In this Project addresses the key Problem:- **How can we accurate predict future sales of product listed on Amazon to optimize inventory levels, enhance marketing strategies and improve overall supply chain efficiency?**

# PROPOSED SOLUTION

- The proposed system aims to address the challenge of predicting the future sales of Products listed on Amazon to improve overall supply chain efficiency. This involves leveraging data analytics and machine learning techniques to forecast demand patterns accurately. The solution will consist of the following components:
- Data Collection:
  - Gather historical Sales data , Product information , Customer Information, including time, date, Region , and other relevant factors.
  - Utilize Product price, discount, Customer purchase history, to enhance prediction accuracy.
- Data Preprocessing:
  - Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.
  - Feature engineering to extract relevant features from the data that might impact sales forecasting.
- Machine Learning Algorithm:
  - Implement a machine learning algorithm, such as a Linear Regression, Random Forest and Decision Tree model to predict sales of product based on historical patterns.
  - Consider incorporating other factors like Holiday, shopping events and seasonal trends ,day of the week, and special events to improve prediction accuracy.
- Deployment:
  - Develop a user-friendly interface or application that provides real-time predictions for sales of the product different time zone.
  - Deploy the solution on a scalable and reliable platform, considering factors like server infrastructure, response time, and user accessibility.
- Evaluation:
  - Assess the model's performance using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or other relevant metrics.
  - Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.
  - Result: 81 % Accuracy rate of the model

# SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the Sales Prediction of Products. Here's a suggested structure for this section:

- Data Collection :- Use APIs, web scraping
- Data Preprocessing :- Python Library (pandas, numpy)
- Exploratory Data Analysis :- (matplotlib, seaborn and plotly)
- Model Building :- (scikit-learn, linear regression, random forest, decision tree, gradient boosting )
- Model Evaluation and Selection :- (MAE, MSE, MAPE)
- Model Deployment :- (Azure, AWS)
- Reporting and Visualization :- (Power BI, Tableau)

# ALGORITHM & DEPLOYMENT

- In the Algorithm section, describe the machine learning algorithm chosen for predicting sales. Here's an example structure for this section:
- **Algorithm Selection:**
  - By leveraging of gradient boosting algorithms, the Amazon Sales Data Prediction project can achieve high accuracy and provide valuable insight into sales patterns and trends, ultimately supporting better business decision.
- **Data Input:**
  - Gather historical data, product information, promotional data, customer demographics and external factor.
- **Training Process:**
  - Import gradient boosting machine by use of appropriate libraries as scikit-learn. Perform hyperparameter tuning to find the optimal setting as number of estimators, max depth, min chills weight, subsample and colsample\_bytree.
- **Prediction Process:**
  - Use the trained model to make predictions on the preprocessed new data.
  - Prediction = model.predict (x-test data)

---

# RESULT

The Amazon Sales Prediction project can effectively use Gradient Boosting Machine Algorithms to generate accurate sales predictions. Thereby supporting better inventory management, marketing strategies and overall business operation.

# CONCLUSION

- The Amazon Sales Data Prediction project highlights the power and potential of Gradient Boosting Machine algorithms in predicting complex sales patterns.
- By leveraging advanced machine learning techniques, the project provides significant business value, enabling better resource allocation, optimizing marketing efforts, and enhanced overall strategic planning.
- As the project evolves incorporating more data and refining models will continue to drive improvements and maintain a competitive edge in the dynamic retail market.



# FUTURE SCOPE

- Continuous improvement of algorithms using advanced machine learning techniques can lead to more accurate sales predictions.
- Implementing real-time data processing and prediction models can help in making immediate business decision making.
- Integration with External Data Sources can enrich the predictive models and provide a more holistic view of market trends.
- Creating sophisticated visualization tools and dashboards to present sales predictions and insights in an easily understandable and actionable format for stakeholders.

# REFERENCES

- [Sales Forecast for Amazon Sales with Time Series Modeling | IEEE Conference Publication | IEEE Xplore](#)
- [GitHub - SayamAlt/Amazon-Sales-Data-Analysis: In this project, I have analyzed Amazon sales records, defined KPIs\(Key Performance Indicators\) and established meaningful relationships between them for deriving useful statistical insights.](#)
- [Amazon Sales Data Analysis - Notebook by Jeef Rojas \(je9rojas\) | Jovian](#)
- [How to use data analysis for machine learning \(example, part 1\) - Sharp Sight \(sharpsightlabs.com\)](#)
- [8 Machine Learning Models Explained in 20 Minutes | DataCamp](#)
- [What is Predictive Analytics? | IBM](#)

# COURSE CERTIFICATE 1

In recognition of the commitment to achieve  
professional excellence



Deepak Behera

Has successfully satisfied the requirements for:

Machine Learning for Data Science Projects



Issued on: 15 JUL 2024

Issued by IBM

Verify: <https://www.credly.com/go/37MwhUfj>



# COURSE CERTIFICATE 2

In recognition of the commitment to achieve  
professional excellence



Deepak Behera

Has successfully satisfied the requirements for:

Data Fundamentals



Issued on: 28 JUL 2024

Issued by IBM

Verify: <https://www.credly.com/go/v3b8unGL>





**THANK YOU**