

# Machine Learning | Statistics | NLP | Generative AI | Data Structure



**Priyabrata Sahoo**  
Data Scientist@ eclerx

## Contact Details:

☎: 9438012606/8260060605

✉: [Priyabratlinkedin@gmail.com](mailto:Priyabratlinkedin@gmail.com)

[Priyabratasahoo55@gmail.com](mailto:Priyabratasahoo55@gmail.com)  
[www.linkedin.com/in/priyabrata-sahoo-a8693b18b](https://www.linkedin.com/in/priyabrata-sahoo-a8693b18b)

## Career Overview:

To understand the business use cases from clients and convert them into a well-defined problem statement and explain it to the development team.

- **6.4 Years in IT Industry, of which 4.5+ years in Data Sciences/ Data Analytics** having worked with MNC's like **Eclerx, Morningstar, Capgemini** across technologies like **/Machine Learning/Deep Learning/ NLP/Computer Vision**.
- To identify data sets required to develop predictive models for solving internal and external business problems.
- To fill data gap by gathering data, designing annotation portal and conducting data annotation by human annotators.
- To explore data sets and identify data transformation and data quality needs for targeted applications.
- To develop algorithms and predictive models to derive insights and business value from data.

## Technical Skillset:

<b>Data Science Stack</b>	<b>Supervised Learning:</b> Good knowledge of Various Machine learning Algorithms like Regression techniques, Decision trees, Random Forest, Naive Bayes, KNN, SVM models, Boosting- XG-Boost, Ada-Boost. <b>Unsupervised Learning:</b> Clustering techniques like Hierarchical, K means, DBSCAN, PCA, Collaborative filtering, Market Basket Analysis (MBA), Association Rules Mining, Customer Analytics <b>Anomaly Detection</b> problems using Isolation Forest, DBSCAN, LOF, OneClassSVM, Statistical Modelling <b>Deep Learning:</b> Neural networks like ANN, CNN, RNN, LSTM, Auto-Encoders, Deep Learning Frameworks like TensorFlow, Keras,Bert, Transformer <b>Time series Forecasting:</b> Smoothing techniques, ARIIMA models. <b>Text Mining, NLP, Sentiment Analysis, Text Summarization, NER (NLTK, Spacy, BERT,GENERATIVE AI, LLM)</b>
<b>Hadoop &amp; Big data Stack</b>	<b>Good Overview of Big Data technologies</b> like HDFS, Hive, Spark, <b>PySpark</b> NOSQL Databases like MONGO DB Deployment Tools: <b>Version Control: Bit-Bucket, Git,AWS,GCP</b>
<b>Predictive analytics using</b>	<b>Python</b>
<b>Data Visualization using</b>	Tableau, ggplot in R and Matplotlib, Seaborn, Plots in Python, MS-Excel
<b>Other technologies</b>	Strong Knowledge of Core Java

<b>Statistics</b>	Descriptive and Inferential Statistics, Hypothesis testing, Anova Regression techniques like linear, logistic, multinomial logistic, Lasso, Ridge, Imputation techniques
-------------------	---

**Education:**

Course	Institute/University	Marks	Year
BSC	Swami Vivekananda Memorial	60%	2016
12 <sup>th</sup>	Royal Science College	71%	2013
10 <sup>th</sup>	Jagannath Uchha Vidya Pitha	65 %	2011

**Certifications:**

- Certified in master's in data science from Simplilearn
- Certified in Master in Machine learning from ineuron.

**Professional Experience:**

From-To	Company	Designation
Mar 2023 to August 2023	ZYMR	Senior Data Scientist
November 2021 to January 2023	Eclerx	Process Manager(Data Scientist)
Apr 8, 2019–June 2021	Morningstar	Machine Learning Engineer(Data Scientist)
Jan 17, 2017- Apr 4 2019	Capgemini	Software Engineer(Data Analyst)

**WORK EXPERIENCE:****ZYMR:-****Sales Forecasting-(Time Series,Regression Model,Databricks,pyspark,Azure AI ,CICD pipeline,MLFlow,Unstructured Learning)**

**1-1010data :-** We have created model based on the store provided to us so initially we first cluster the product into 3 categories like fast movers /average movers, slow movers and outliers and based on the clustering we have worked on forecasting on different products.

- Construct Time series model on Arima model.
- Constructed the Regression Model and worked on error analysis for different errors like MAE,MSE.
- I have used Databricks and ml flow for end-to-end project deliverables.

**eclerx****NOV 2021-Jan 2023****1-Weekly and Monthly Sales prediction of GANNETT (Time Series Modelling)**

We created model for Gannett client with their specific requirement on time series .They have different site for which they required their forecast to build so for weekly we used to create model and run on every week and send them the report and based on that report they see their sales growth and monthly basis as well we used to make model and send them report on monthly basis for their result.

- Construct Time series model on Holt Winter and fbp rophet model
- I have made the Holt Winters model and fbprophet model for weekly data.

### **Technology**

Regression, Classification, Python, GCP,LSTM

Involved in requirement gathering and architecture design.

- Building data cleansing pipeline to train a model.
- Evaluated model with RMSE and MSE
- Developing time series model for the train data to build model
- Presented to the Client with proper documentation.

### **2-TEXT SUMMERISATION USING NLP (USING LLM MODEL AND GENERATIVE AI AND TRANSFORMER)**

We have created a model which will summarize the document by using LLM model and Generative AI technology.

. We made our own Customized LLM model for Text classification and to detect the object we used Transformer techniques object.

- Construct LLM Architecture using Transformer Technologies
- I have made the UI and integrated the same with UI using flask.
- Created the LLM Model with Transformer and used Latest Technologies.

### **Technology**

Deep Learning, Computer Vision, Python

Involved in requirement gathering and architecture design.

- Building data cleansing pipeline to train a model.
- Developing object detection models for single grains that is wheat image classification.
- Presented to the Client with proper documentation.

## **Morningstar**

**Apr 2019-June 2021**

### **1-Commercial Paper Bond- Financial Domain (Using Random Forest and XgBoost)**

Commercial Paper is a short-term bond where there is not many risks is associated and whenever a company wants to raise money for a short term so they issue these short terms bond so they will take a loan and they will issue these bonds and investors can invest in these bonds.

### **Technology**

Machine Learning, Random Forest ,GCP

### **Contribution**

Involved in requirement gathering and architecture design.

- Building data cleansing pipeline to train a model.
- Developing statistical models for various predictive methods such as forecasting, classification, clustering, and regression.
- Involved in parameter tuning process for optimal model hyperparameters.

### **2-Word Document Checking using Sentiment Analysis (Bert, sentiment analysis, NER)**

We as Morning Team collected the 10k file and we analyzed the report based on the risk factor that each company has and we predicted using sentiment analysis on the risk part on 10k file which is available. Every company published this report so that investor who wants to invest in their company they should aware wheatear their investment is safe or not so we as Morningstar we have created a sentiment analysis which will help Us in predicting about the company growth.

**Technology: -**

NLP (Sentiment Analysis on 10k filing)

**Contribution: -**

Involved in requirement gathering and architecture design.

- Building data cleansing pipeline to train a model.
- Developed sentiment analysis model on the given dataset.
- Involved in parameter tuning process for optimal model hyperparameters.

**3-Funds Growth Rate:- Financial Domain (ARIMA,LSTM)**

Provide fund growth rate factor exposures across the equity, fixed income, and allocation asset classes.

- Construct the monthly growth rate return associated with each factor.
- Forecast fund flow growth rates up to one year in the future.

**Technology**

Machine Learning, AWS,LSTM,RNN

**Contribution**

Involved in requirement gathering and architecture design.

- Building data cleansing pipeline to train a model.
- Developing statistical models for various predictive methods such as forecasting, classification, clustering, and regression.
- Involved in parameter tuning process for optimal model hyperparameters.

**Capgemini**

Jan 2017-April 2019

**Member' world claim Portal for Bupa Client: - Insurance Domains**

Member's world claim application used to receive multiple claims and on basis of parameters we need to check the claim is whether valid or not and we check whether the claim has been fraudulent or not and claim approval process.

**Technology**

Python, Data Analysis

**Contribution**

- Involved in requirement gathering and the concept design phase of the project.
- Developing statistical models for various predictive methods such as forecasting, classification, clustering, and regression.