

Prabhat Chandra Shukla

 Bhopal(M.P.), INDIA
 +91 99817-83437

prabhu_shukla19@hotmail.com

<https://www.linkedin.com/in/prabhatcs>

A dedicated data science professional having 12+ years of overall experience with 8+ years of relevant experience in data visualization, machine learning, NLP, predictive modelling, and statistical modelling. Able to play key role in analyzing business problems and come up with creative solutions.

Skills:

- **Area(s)of Interest** Generative AI, Machine Learning, NLP, Neural Network, Data Mining, Chatbot, Data Visualization, Image Processing, Deep Learning.
- **Programing Language** Python, PySpark, R
- **Database** MySQL, Neo4j, Big Query, DB2, Rds
- **Cloud Platform** Google Cloud, Azure, Data Bricks, AWS
- **Tools and Frameworks** Power BI, Power Apps, Streamlit, Spark, Long Chain, LLM

Education:

Course	Institute	Board	Year	Percentage
PGD in Big Data Analytics	Siddaganga Institute of Technology, Tumkur	Autonomous Affiliated to VTU	2016-2017	8.69 CGPA
B.E. (CS)	RKDF College of Engineering Bhopal	Rajeev Gandhi Technical University	2007-2011	67.34%

Certifications:

C3.AI V8 DATA SCIENCE	C3 Ai Academy
Deep Learning Specialization	deeplearning.ai (Coursera) [Andrew Ng]
Machine Learning with Big Data	University of California, San Diego (Coursera)
Graph Analytics for Big Data	University of California, San Diego (Coursera)
Machine Learning Foundations	University of Washington (Coursera)

EXPERIENCE

Yash Technologies Pvt Ltd, Indore — Sr. Data Scientist

Feb 2020 – Present

Client: John Deere (Current)

Projects:

❖ **Automated Warranty Responsibility Code Assignment**

Project description: Automatically assigning responsibility codes to warranty claims, determining whether the supplier or Deere is liable, based on detailed complaint descriptions and associated cause and correction texts.

ML Problem: Text Classification, Textual Data Analysis

Technologies: Random Forest, BERT, LLM, vLLM

Data: EDL, Databricks

❖ **Enhancing Extended Warranty Data in Palantir Foundry**

Project description: To improve the quality, accessibility, and actionable insights derived from extended warranty data by leveraging Palantir Foundry's data ingestion, enrichment, and analysis capabilities.

Problem: Data Ingestion(ETL), Data Enrichment, Features Enhancement

Technologies: Palantir Foundry, PySpark,

Data: Rds, EDL

❖ **Attachments Parts Forecasting (Time Series)**

Project description: Design a forecasting mechanism for attachments parts of finish goods machine, by using historical sales of finish goods as well as attachments parts. Evaluate model based on key parameters and measures and develop insight of outcome.

ML Problem: Data Visualization, Time series prediction

Technologies: Prophet, ARIMA, Holt-Winter, Pandas, Plotly

Data: Hana, Db2

❖ **Other Projects**

1. LLM Based PDF Summarizer using DeereAI, LongChain.
2. vLLM based feedback transcripts analyzer.
3. Reports automation using Power BI and Power Apps
4. Building automated data collection and cleaning pipeline using Databricks.

Technologies/tools: Python, R, DeereAI/OpenAI, vLLM, PowerBI, Tableau, Databricks, Power Apps/ Automate

Client: PALL Corporation

Projects:

❖ **Flowstar (Filter Integrity Test Instrument) Visual & Predictive Analytics**

Project description: Pall offers a Flowstar machine to their pharmaceutical customers to test the integrity of the filter. They are required to do this test due to regulatory requirements. Testing integrity means, testing to make sure the filter is fit enough to perform its function, a very important criterion when it comes to medicinal drug manufacturing.

Purpose of this project was to

- I. Analysis of the integrity test results and develop reporting analytics around it.
- II. Develop a machine learning model to predict the probability of result on early stage

Role:

- Created pipeline that collect all log files (local directory) which was in XML format,
- Perform data cleaning, feature engineering steps on raw data and create visual reporting out of it.
- Trained machine learning model to predict the probability of result on early stage using previous log data.

ML Problem: Data Visualization, Classification

Technologies: scikit-learn, plotly, seaborn, pandas

Data: Log data as XML format

❖ **Sales Forecasting (Time Series)**

Project description: Purpose of this project was to Develop Sales forecast model for top Product Families/SKUs in the Food & Beverage Market. Develop insight for Product families and Evaluate model based on key parameters and measures.

Role:

- Collecting dataset from database
- Data cleaning and initial reporting(EDA)
- Create a pipeline for try and validate different time series models and select best one based on less error rate.
- Create report based on forecasting data that contains comparing for different time frame as well as product families.

ML Problem: Data Visualization, Time series prediction

Technologies: Prophet, ARIMA, Holt-Winter, Seaborn, Pandas, Plotly

Data: SQL Database

Accenture Solutions Pvt Ltd, Gurgaon — Sr. Data Scientist

July 2018 – Feb 2020

Client: Google, Inc.

- Leverage a wide range of data analysis, machine learning and statistical modelling algorithms and methods to solve business problems in *Google Retail*.
- Generate deep insights through the analysis of data and understanding of operational processes and turn them into actionable recommendations.
- Build and deploy prototype solutions to demonstrate ideas and prove concepts.
- Created automated solutions for client by utilizing Google Cloud applications and python shell scripting.
- Generated data driven and analytical insights to be presented to client and stakeholders for effective decision making.

Projects:

❖ Market Basket Analysis

Project description: Purpose of this analysis is to generate a set of rules that link two or more products together. We are interested in the support and confidence of those rules: higher confidence rules are ones where there is a higher probability of items on the RHS being part of the transaction given the presence of items on the LHS.

ML Problem: Recommendation system (Association rules mining).

Algorithms: Apriori algorithm

Implementation: Scheduled execution on Colab and export to Big Query

❖ Email Sent Time Optimization

Project description: To make email campaigns more impactful, client wanted to predict the best time to send emails to the users so that open and click rate can be increased. To predict this, I analyzed user behavioral trends for a certain time period email data and predicted best time to send emails to the users for effective engagement.

ML Problem: Regression, Classification

Algorithms: Logistic Regression (Multinomial), Decision Tree, Random Forest

Implementation: Colab

❖ Segmentation of Site Visitors using Spark

Project description: Client wants to create different segment of the visitor's data that visiting store site to see what all clusters/segments get formulated if data is directly fed into the ML model. Since they had created manual logic to differentiate user types, namely: Buyer, Shopper, Browser and Bot traffic.

ML Problem: Clustering

Algorithms: K- means Clustering

Implementation: PySpark

❖ Other Projects

1. Customer Acquisition and Retention Analysis (CLV).
2. Propensity Modelling.
3. iTunes data scraping and manipulation.
4. Weekly Site Session Forecasting.

Northout Solutions, Indore — *Data Scientist*

Dec 2017 - July 2018

Project:

❖ Spending Habits Analysis

Project description: Using transaction data of user bank account/credit card we need to find spending habit and transactional behavior of user. We also have additional details such as demographic, investment, family, educational, career related information, lifestyle and major life events like birthday, house change, marriage anniversary. Using all the above provided details we need to create visualization and spending trends based on user's additional details and major life events.

Client: John Hancock Financial

Role:

- Collecting dataset from plaid.
- Create Visualization on user transactions per category.
- Identify spending habits by major life events
- Predict user transaction for upcoming months.

ML Problem: Data Visualization, Time series prediction

Algorithms: ARIMA, regression

Technologies: scikit-learn, prophet, scrapy, keras, seaborn, pandas

Data: Plaid API, Credit card transaction details

Implementation: Web App

Bonsmat Group, Ludhiana — Machine Learning Engineer

Sept 2017 - Nov 2017

Project:

❖ **ChatBot Assistant:**

Project Description: Idea was to develop conversational assistant dialogue system, which can provide latest offer related to mobile recharge using Rest API and also able to do recharges of mobile numbers. Having features of weather forecasting, news summary and question answering system (using Wikipedia API).

Client: bgpay.in

Role:

- Collection of the News article.
- Create Machine Learning model for text classification.
- Create Machine Learning Model for NER.
- Create Rule based model to identify the mobile operator name, amount and circle.
- Create API system to interact with client server.

ML Problem: ChatBot, Classification

Algorithms: Named Entity Recognition, Intent recognition, Text summarization

Technologies: NLTK, AIML, ChatterBot, Spacy, Beautiful Soup, Flask.

Data: News feed, Wikipedia, user input text.

Implementation: Web App, Chrome Extension

Constalytics, Mohali — Data Engineer

April 2017 – Aug 2017

Project:

❖ **Knowledge Graph Platform:**

Project Description: Unstructured text data processing platform which helps to analyze the text using name entity extraction, topic modeling and sentiment analysis. This platform helps to create knowledge graph using Neo4j to extract possible relationships between entities.

Role:

- Unstructured data sources integration.
- ML model for text classification and Name Entity Recognition.
- Integration of Neo4j for knowledge graph creation.

ML Problem: Natural Language Processing, Classification

Algorithms: Named Entity Recognition, Intent recognition, Text summarization

Technologies: Spacy, Beautiful Soup, Flask, Neo4j

Data: Unstructured text data.

Implementation: Web App

Data Science Research Institute, Bengaluru — Machine Learning Researcher

Aug 2016– March 2017

Projects:

❖ **Cluster Analysis Using Spark:**

The aim of the project was to find the change in the weather for California State from year 2011-2014. Here I used clustering technique to perform analysis and find some interesting patterns in the dataset (weather.csv) and later cluster them using Kmeans algorithm.

❖ **Cricket Prediction and Analysis:**

Scrape the details of cricket players from “www.espnricinfo.com” and based on the previous records, their performance can be analyzed for future. Here I analyzed player performance and developed prediction model using *CricketR* package in R.

Predictive Research, Bengaluru — Software Developer

Mar 2015– Aug 2016

Freelancer, Bhopal — Web Developer

Jan 2012 – Feb 2015

Declaration:

I hereby declare the above details are correct to the best of my knowledge and belief.

Place: Bhopal

(Prabhat Chandra Shukla)