

POOJA SINGH

MASTER'S IN DATA SCIENCE | BACHELOR IN COMPUTER SCIENCE | Developer | PREDICTIVE ANALYTICS |

DATA ENGINEERING | Machine Learning | DEEP LEARNING | NLP | Gen AI

 FURLONG, PA | 4454485034  MAIL.POOJASINGH11@GMAIL.COM

Experienced Python Developer proficient in cloud-based application design, database optimization, and machine learning integration. Skilled in data fetching and cleaning, API development, web crawling, HTML, CSS and JavaScript. Strong collaborator with excellent troubleshooting abilities, dedicated to delivering smooth user experiences and safeguarding data integrity. Active participant in the Python open-source community, passionate about ML and GENAI, and committed to continuous learning and staying updated on the latest data science technologies.

CORE COMPETENCIES

ML Algorithms and Modeling: Supervised Learning, Unsupervised Learning, Deep Learning, AutoML, stacked ensemble models, Clustering Techniques Logistic Regressions, Linear Regression, Decision Trees, Naïve Bayes, NLP (NLTK, Spark NLP, spacy), NLP Classifier, KNN Classifier, Sentiment Analysis, Deep Neural Network (DNN), Convolutional Neural Network (CNN), n/w Flow Models (Transshipment and Max Flow Problems), Oversampling, Feature Engineering, Descriptive Statistics, Visualization, and Feature Selection (Embedded & Baruta).

Databases and Engineering: Data warehousing, ETL, Database Management, My SQL, S3, a basic understanding of RDBMS tools, including SQL, SQLITE3, Oracle, Excel VBA

Environment: Windows, Linux, Git, Django, Fast API, Python, R, MYSQL, MS SQL Server, Oracle, Anaconda, Colab, Pycharm, Visual Studio, AWS S3.

Cloud: AWS S3, AWS Glue, AWS Data Warehouse

Visualization: Tableau, Matplotlib

Generative AI Techniques: Retrieval Augmented Generation, Finetuning techniques (PEFT, LoRA, QLoRA), Prompt Engineering, Chatbots, Langchain, Hugging Face Hub, Vector Databases Chromadb, pinecone, FAISS

Large Language Models: Llama2, OpenAI, ChatGPT, BERT, RNN, LSTM etc

Languages and packages: Python, HTML, CSS, SVG, JavaScript, D3.js; predictive modeling utilizing various Python packages, including Pandas, NumPy, Scikit learn, Seaborn, SciPy, Matplotlib; TensorFlow, Pytorch, Keras

Microsoft Office suite: Word, Excel, PowerPoint, Outlook

EDUCATION

New Jersey City University, Jersey City, NJ:

GPA: 4.0

Master of Science in Business Analytics and Data Science, 2023

Gyanchandji Hirachandji Raison College of Engineering, Nagpur:

Bachelor of Engineering (B.Eng.) in Computer Science, 2010

EXPERIENCE

Inclusion Bridge

July17-August11, 2023

- Conducted an in-depth analysis in the Commercialization IT Data Capabilities division, assessing the impact of ESG data on the financial client's "Predictive Trade Model". Evaluated input factors (spread, liquidity, volatility) and output metrics (trade impact).
- Utilized a sophisticated array of correlation analysis methods, including both linear and non-linear approaches such as Pearson Correlation and Distance Correlation, alongside the computation of correlation P-values. These techniques unveiled intricate relationships between ESG components and trade impact, empowering

data-driven trading decisions. Investigated "Absolute Distance Correlation Differences" for evolving dependencies. Applied statistical validation (P-values) for significance.

Environment: Python(Numpy, Pandas, Seaborn, Matplotlib, Plotly, Statsmodels, A/B testing, Scipy, dcor(distance correlation), Scikit-learn, Jupyter, Tableau, Statistical validation(P-values), Hypothesis Testing, Data Visualisation, Trade Impact Analysis.

Steel Authority of India Limited, Bokaro, India,

Sept- 2022 to Nov- 2022

Role: Predictive analytics on Inventory management Internship

- Applied statistical models and data analysis for precise inventory forecasting and management optimization, minimizing waste.
- Led the creation of an advanced predictive inventory model at Steel Authority of India Limited, transforming inventory management. Converted the model into a versatile training tool for scenario risk modeling.
- Introduced an efficient, data-driven methodology, revolutionizing inventory practices. Established the model as an asset for proactive risk mitigation and strategic planning.

Environment: Python(Pandas, seaborn, Predictive Models, Numpy, Scipy, Sklearn(KFold, cross_val_score, train_test_split, GridSearchCV, mean_squared_error, r2_score, mean_absolute_error), Data Visualisation)

Quantum

July -2021

Role: Virtual program Intern

- Interned at Quantum as a Virtual Program Intern mainly focused on Data Preparation and Customer Analytics, using experimentation and uplift testing analytics to uncover practical insights

CASA Indian Homes Private Ltd (Real Estate and Infrastructure)

Sep 2014 – June 2019

Role: Data Engineer and Analyst (Team member of strategy and marketing team working for real estate Demand Forecasting and Pricing Data)

Responsibility:

Data Warehousing, Data collection, Data cleaning, Data transformation and loading data on AWS AWS-based MySQL server. Real Estate Data collection and extraction from the Web (Web Scrapping), online property sites, API, Data Cleaning, preparation and Transformation(AWS Glue and Lambda), and storage (My SQL). Geographical, demographical, historical data, Financial Data

Market research Data Management, Data Modeling for real estate demand forecasting, and pricing decision

Projects:

- New Construction price prediction modeling for the next five years of demand- Creating a predictive analytics model for real estate price forecasting.
- Developed a customer-centric demand forecasting model in Python, using precise customer segmentation to optimize marketing and resource allocation strategies, thus enhancing overall business performance and supply chain efficiency for a real estate company.

Environment: AWS-based environment utilizing RDS MySQL, S3, Glue, Lambda, Python (Pandas, Scikit-learn), Jupyter, Git.

ACADEMIC PROJECTS AND PAPERS

- **Twitter NFT Sentiment Analysis:** Earned distinction with an accepted and published paper at the esteemed 49th Annual Northeast Business and Economics Association Conference, Employing API and web scraping techniques for meticulous data collection, adeptly addressing API limitations, centered around Python-driven sentiment analysis on Non-Fungible Tokens.
- **Multi Weather Classification (Deep Learning), Fall 2022:** Successfully classified multi-weather conditions using Convolutional Neural Networks (CNN) and transfer learning, elevating test data accuracy from approximately 22% to approximately 81%.
- **Business Analyst Job Listing Analysis, fall 2022:** Engineered and executed a web-based interactive visualization and comprehensive report utilizing d3.js, HTML, and CSS.
- **Case Study, Fall 2022:** Orchestrated a comprehensive case study resolving network flow challenges for Bruce McDaniel at Dominion Energy, Inc. Leveraged Linear Programming (LP), Solver, and other tools to devise a solution that tracked gas availability across nodes, minimized transportation costs, and maximized profits.

- **Student Dropout Prediction Challenge (3rd position in Kaggle Challenge), Spring 2022(course: Machine Learning)** Evaluated classification techniques—Bagging, SVM, Random Forest, Stacking, and Logistic Regression—on NJCU students, constructing predictive models (R) for gauging the likelihood of student attrition. Transformed analyses into impeccably documented reports through R markdown.
- **Patient Procedure Information System for Hospital, Spring 2022(course: Developing Financial Software)** Spearheaded the development of a menu-driven, object-oriented project for patient data management. Employed SQLite3 for SQL-Python integration and Inheritance for medical procedure cataloging.
- **Human Resource (HR) Employee Management System (course: Data Management), Spring 2022:** Created an encompassing HR database in MySQL for streamlined employee information management.