Pavan Teja Nagisetti
Phone : 267-298-0143
Email : ptn36@drexel.edu

- Extensive years of experience as a Machine Learning & Data Scientist, including profound expertise and experience in statistical data analysis such as transforming business requirements into analytical models, designing AI algorithms, and strategic AI solutions that scale across massive volumes of data.
- Good knowledge of Agile methodology and Waterfall methodologies ability to manage all phases of SDLC ranging from requirement analysis, design, development, testing to the deployment of a data science project.
- Used Pandas, NumPy, Scikit-learn in Python for developing various machine learning models such as Random forest and stepwise regression and a good working knowledge of Python libraries including Ggplot2, NumPy, Pandas, SciPy Matplotlib, Seaborn, and Scikit – Learn for analysis purposes.
- Regarded as an inventor at Mastercard, for developing two AI systems which can identify information into a desired class based on Gen AI, Ollama, Langchain, NLP, pattern mining and decision trees. Developed deep learning models such as Image Recognition, Object Detection, Facial Recognition with an accuracy greater than 93% using transfer learning.
- Developed and implemented AI models using LLMs and also traditional models such as linear regression, classification, multivariate regression, Naive Bayes, Random Forests, K-means clustering, and KNN.
- Hands-on experience training, developing and deploying models over EC2 Instances.
- Well-versed in database querying using MySQL, PostgresQL, and Snowflake. Experienced in ETL technologies such as Talend and Alteryx.
- Strong roots in Business Intelligence and Dashboarding using multiple tools such as Tableau, PowerBI, Google Data Studio, ChartIO, and Vizard.

EDUCATION

Master of Science in Business Analytics,
Lebow College of Business, Drexel University   Philadelphia, PA
Bachelor of Technology in Mechanical Engineering,
Indian Institute of Technology Guwahati

                                                                                      Guwahati, AS

TECHNICAL SKILLS

Pogramming Skills: Python,Flask, Excel, R, SAS,
Data Engineering : Talend (ETL), Alteryx, Snowflake, Postgresql, MySQL, MsSQL,
Data Visulaisation : Tableau, Power BI, Google Analytics,
Cloud Exposure : Azure, AWS, EC2, ssh, sftp,
Project Management & Version Control : JIRA, Confluence, Github, Pug,
Data Science Skills : Linear Algebra, Calculus, Data Analytics, Algorithms, Regression, PCA, Factor Analysis, Decision Trees, Random Forest, SVM, Naive Bayes, KNN, Kmeans clustering, Sentiment Analysis, Sequence Models - RNN, LSTM, Convolutional Networks - Siamese Networks, Keras, Tensorflow, Langchain

WORK EXPERIENCE

MasterCard, Salt Lake City, UT   **Oct 2021 – Present**
ML Engineer / Data Scientist

Regarded as an Inventor for patenting two AI systems which can identify information into a desired class based on NLP, pattern mining and decision trees.

Responsibilities:
Transaction Categorization – NLP:
● Developed and implemented a cutting-edge Smart Labelling Framework aimed at semi-automating transaction categorization processes. Achieved an impressive accuracy rate of 89.3%, significantly reducing manual effort and improving operational efficiency.
● Engineered a comprehensive product leveraging Natural Language Processing (NLP) techniques and rule-based algorithms to categorise transactions into 120 distinct categories. This solution was developed on a vast dataset comprising 125 million transactions, enhancing transaction tracking and analysis capabilities.

Personal Information Redaction – NER:
● Pioneered the creation of a groundbreaking Personal Information Identification, Extraction, and Redaction model (PIER) from inception using Gen AI and traditional NLP based Entity recognition engines..
● Iteratively improved the model, achieving a recall rate of 95% in the MVP version. Subsequent versions demonstrated precision rates of 72% and recall rates of 86%, showcasing continuous refinement and optimization efforts.
● Successfully realised the product's deployment, enabling the anonymization of personal information within unstructured columns of a vast database (~3 billion transactions) with an F1 score of 78%.
● Acted as the primary point of contact and collaborated closely with cross-functional teams, including Human Labellers, Data Engineering, and Product Owners, to ensure seamless integration and adoption.

Rent Payment Identification:
● Engineered an automated pipeline to accurately identify residential rent transactions for Experian, a leading credit bureau.
● Delivered the algorithm with an outstanding accuracy rate of 87%, facilitating precise categorization and analysis of rental-related financial data.
● Developed a sophisticated smart labelling framework that autonomously labelled 97% of non-rental data within a dataset of 8 million rows, streamlining data preprocessing and enhancing overall efficiency.

Cashflow & Balance Analytics for Small and Medium Enterprises:
● Served as a Technical Lead, spearheading the creation of a comprehensive framework for product realisation. Led all aspects of the project lifecycle, including requirement gathering, technical development, and automated pipeline implementation for testing and edge cases.

PATH, Seattle, WA   **Jul 2020 – Dec 2020**
Machine Learning Engineer/Data Scientist

Responsibilities:
● Designed and implemented multiple dashboards in Tableau using parameters, filters, and calculated fields to visualise 150+ healthcare
● policies, 26 categories, 76 countries across the world for WHO and The Gates Foundation.
● Integrated data from multiple data sources using Alteryx to clean, organise and consolidate the data for visualisation in Tableau.

Revzilla, Philadelphia, PA   **Jun 2020 – Sep 2020**

Data Scientist

Responsibilities:
- Developed an ML model in Snowflake with advanced SQL functions and CTMs to improve performance accuracy of customer segmentation/cluster models on a data set of around 8 million rows extracted from Google Analytics from 53% to 74 %.
- Worked on model development, scoring, and validation scripts programmed in Python and Snowflake for digital analytics to understand customer loyalty and non-returning customer behaviour based on their visit(s) and purchase patterns.

Redeem Systems Consulting, Bangalore, IN   **Aug 2018 – Jul 2019**
Data Scientist - Machine Learning & AI

Responsibilities:
- Recognised as an SME in Data Science and AI, trained and led a team of 8 data analyst(s) to execute the build, test, delivery cycle of 26 machine learning projects.
- Hands-on experience in developing Machine Learning, NLP, Customer Analytics, Risk Modelling, Language Models, and Deep Learning models in Sklearn, Keras, and Tensorflow while interfacing with cloud solutions provided by Microsoft Azure's Cognitive Services.
- Created Image recognition, Object Detection, Facial Recognition models with an accuracy greater than 95 % using transfer learning. Packaged Machine Learning Algorithms by developing REST API(s) and deployed models to production using the Python libraries Sklearn, Pandas, Flask, and Amazon EC2 instances. Used GitHub for version control and code collaboration with team members.

Redeem Systems, Bangalore, IN   **Apr 2018 - Aug 2018**
Data Science Associate

Responsibilities:
- Improved the performance accuracy of Machine Learning and NLU based Classification Models (Chatbot) from 87 to 93% using Sklearn and Pandas. Hands-on experience in deploying ML models as APIs using ssh and sftp protocols over AWS EC2.
- Used Python (NumPy, SciPy, Pandas, Scikit-Learn, Seaborn to develop a variety of models and algorithms for analytic purposes.
- Developed multiple Dashboards for reporting and monitoring KPIs for a chatbot using the data visualisation tool ChartIO.
- Determined trends and relationships in data by applying advanced statistical methods like T-test, hypothesis testing, ANOVA, Chi-Square test, and Correlation Analysis.

Infruid Labs, Hyderabad, IN   **Nov 2017 - Feb 2018**
Data Analyst(Business Intelligence)

Responsibilities:
- Responsible for creating and maintaining client dashboards across verticals such as Manufacturing and Healthcare using the BI tool Vizard to monitor KPIs.
- Experienced in using Talend for Data Ingestion(ETL), SQL, and Elasticsearch for Database Management, Postman for API Testing.
- Created the documentation and user guide for a Business Intelligence tool - Vizard using the language pug.
- Acquainted with all aspects of Software Development Lifecycle (SDLC) from requirement analysis, Design, Building Coding, Testing, Deployment, & Maintenance Waterfall methodologies.