VENEEL KUMAR A

Bangalore, India | veneeldas77@gmail.com

LinkedIn: linkedin.com/in/veneelkumar | GitHub: github.com/Veneel77

SUMMARY

Aspiring Data Scientist with hands-on experience in machine learning, data preprocessing, and deep learning. Completed the Google Advanced Data Analytics Certificate and built multiple end-to-end projects involving classification, regression, and data visualization. Proficient in Python, SQL, and modern ML libraries. Actively seeking entry-level data science roles to contribute analytical insights and predictive solutions.

TECHNICAL SKILLS

Programming Languages: Python, SQL, Java
Data Science & ML Libraries: NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, TensorFlow, Keras
NLP & Generative AI Tools: Hugging Face Transformers, NLP pipelines (summarization, text classification)
Cloud & MLOps:AWS SageMaker, AWS S3, Boto3, IAM, CloudWatch
Data Visualization & Dashboarding: Power BI, Streamlit
Databases & Querying: MySQL, SQL (advanced querying, joins, subqueries, aggregations) **Version Control & Deployment:**Git, GitHub, Streamlit Sharing, AWS SageMaker Endpoints **Core Concepts:**
Machine Learning, Deep Learning, Natural Language Processing, Model Deployment, Exploratory Data Analysis (EDA), Data Preprocessing, Model Evaluation, API Integration, Web Scraping

CERTIFICATIONS

- Google Advanced Data Analytics Professional Certificate – Coursera (Google) | May 2025
  - Covered data wrangling, statistics, regression, classification, clustering, and data ethics
  - Hands-on with Python, Pandas, scikit-learn, and Tableau.
- Microsoft Azure AI-900: AI & ML Fundamentals
- Data Science for Engineers – NPTEL
- Geo-processing using Python – IIRS, ISRO
- Foundations of DS and ML – Global Training Partner
- Applications of IoT – SST Technologies

PROJECTS

MOBILE PRICE CLASSIFICATION USING SKLEARN ON AWS SAGEMAKER

Tools & Technologies: AWS SageMaker, S3, SKlearn, Jupyter Notebook, Python, CloudWatch, IAM, Boto3

Description:

- Built and deployed a machine learning model to classify mobile phone prices into multiple categories using Scikit-learn.
- Used **AWS S3** for data storage and **AWS SageMaker** for training and deploying the ML model via a **custom training script**.
- Implemented an **end-to-end pipeline**: data preprocessing, training, evaluation, and deployment in a scalable and reproducible environment.
- Managed cloud infrastructure via **IAM roles** and monitored model metrics using **CloudWatch**.
- Automated data ingestion and inference pipeline using **Boto3** and **Jupyter notebooks**

AI-POWERED JOB DISCOVERY AGENT

- Built a Streamlit application that scrapes and filters remote job listings from RemoteOK based on user-defined keywords and tags.
- Integrated open-source transformer models (via HuggingFace) to summarize job descriptions into concise overviews.
- Designed an intuitive UI for interactive job discovery and rapid content skimming.

CUSTOMER CHURN CLASSIFICATION USING ANN

- Built an artificial neural network using Keras to classify banking customer churn
- Preprocessed features including credit score, balance, and loan status
- Achieved 85% accuracy; evaluated model with confusion matrix and ROC-AUC

STUDENT PERFORMANCE PREDICTION

- Developed regression models to predict academic performance based on study habits and demographics
- Trained and compared Linear Regression and Decision Tree models; achieved R^2 of 0.84
- Performed data cleaning, feature engineering, and visualization

NETFLIX DATA ANALYSIS-SQL

- Analyzed Kaggle-based Netflix datasets to uncover trends in content, ratings, and genres
- Wrote complex SQL queries for aggregation, filtering, and reporting insights

DATA PREPROCESSING TOOL (ACADEMIC PROJECT)

- Developed an interactive data preprocessing tool using Python and Streamlit
- Enabled handling of missing data, encoding, scaling, and basic visualizations

OPEN-SOURCE CONTRIBUTION

Open Source Contributor – pandas-dev (Pandas)

- Replaced manual column deduplication logic in CSV parser with shared utility `dedup_names()` to improve consistency
- Submitted PR #61670 to pandas-core (under review)
- Built and tested pandas from source using Cython + Meson (Windows)

EDUCATION

K S SCHOOL OF ENGINEERING AND MANAGEMENT, BENGALURU
B.E in Artificial Intelligence and Data Science Aug 2021 – May 2025 | GPA: 6.5

ST. FRANCIS DESALES PU COLLEGE, TAMAKA
Science (PCMB)
2019 – 2021 | GPA: 7.5/10

S.F.S HIGH SCHOOL, MALUR April 2019 | GPA: 8.1/10

Relevant Coursework: Data Structures & Algorithms, Machine Learning, Deep Learning, Data Science, DBMS, Statistics

WORKSHOPS

- Deep Learning Workshop by Mevi Technologies

ACHIEVEMENTS

- Solved 50+ problems on LeetCode (Data Structures and Algorithms & SQL)

LANGUAGES

Tamil, English, Kannada, Telugu