Pankaj Karki

Github: github.com/PankajKarki

Email: pankaj.karki.786@gmail.com
Mobile: +91-798-2057-772

EXPERIENCE

- **Instoried** — Remote
  Machine Learning Engineer (Full-time) — *May 2019 - Present*
  - **Team management**: As a technical lead, I led a team of ML Engineers and taggers to create impactful NLP solutions like Emotion Classification, Named Entity Recognition, and Part-of-speech tagging with an accuracy of 95%.
  - **NLP**: Utilized Openai APIs to fine-tune GPT3 for various tasks, collected 5 million data points and fine-tuned transformer architecture to train a Bert classification model through transfer learning, and performed behavioural testing, including MFT, Directional, and Invariance tests on NLP models.
  - **Data modeling**: developed and deployed ELT pipelines to ingest data from various sources across platforms and data formats, while also defining and implementing CI/CD best practices using Git, Docker.
  - **Data-centric approach**: By contributing to the team's machine learning practices and collaborating with cross-functional teams, I identified and implemented data-driven solutions to improve the organization's data processing capabilities and reduced data errors by 10%, enhancing data quality.

- **Petpal** — Hyderabad
  Data Analyst (Intern) — *May 2018 - July*
  - **Descriptive analytics**: Utilized Python to build ETL pipeline and dashboards to load data and generate business insights using descriptive statistics and visualization methods.
  - **Predictive analytics**: Analysed and interpreted trends or patterns in complex datasets using statistical and machine learning techniques, such as neural networks, decision trees, and regression.

SKILLS SUMMARY

- **Languages**: Python, JavaScript, SQL, Bash
- **Frameworks**: Numpy, Pandas, Scikit, NLTK, SpaCy, PyTorch, Keras, FastAPI, Django, React
- **MLOps Tools**: Kubernetes, Docker, GIT, PostgreSQL, MySQL, SQLite, DVC, MLflow
- **Platforms**: Linux, Web, AWS (S3, EC2, etc), GCP
- **Miscellaneous**: ETL, Data Science pipeline (cleansing, wrangling, visualization, modelling, interpretation), API load and performance test using Apache JMeter, Postman

PROJECTS

- **Food Reviews Analysis and Modelling(NLP, Machine Learning)**: Performed Exploratory Data Analysis, Data Cleansing, Data Visualization, and Text Featurization(BOW, TFIDF, Word2Vec). Build several ML and DL models like KNN, Naive Bayes, Logistic Regression, SVM, Random Forest, GBDT, LSTM(RNNs), etc.

- **DVC reproducible Pipeline (classification)**: Machine Learning end-to-end reproducible pipeline, logging experiments, versioning data and models.

- **Apparel Recommendation, CNN(Recommendation engine)**: Build a recommendation engine that suggests similar products (apparel) to the given product (apparel) based on features (colour, brand, product description, product Image) in any e-commerce website.

- **Blog-Site application(Backend Project)**: A sample backend application for a blog site was built using fastAPI and SQLite.

EDUCATION

- **Vellore Institute of Technology** — Chennai, India
  Masters of Computer Application — *July 2016 - May 2018*

- **Garden City College** — Bangalore, India
  Bachelor of Computer Application — *July 2013 - May 2016*

CERTIFICATION

- **Machine Learning from Coursera** — Online
  learned about Logistic Regression, Artificial Neural Networks, and Machine Learning (ML) Algorithms. — *Mar 2018*

- **Data Science from DataCamp** — Online
  Completed data science and data analysis track and worked on various projects. — *Jan 2018*