

SANDEEP PANDEY

📍 Kitchener, Ontario, Canada ✉ sandeep.p1296@gmail.com ☎ +1(647)9378396 🌐 spandey.me

EDUCATION

Reporting Systems and Database Development | Conestoga | 3.66 | Kitchener – Downtown | 2023

Bachelor of Software Engineering | Mumbai University | Mumbai, India | 2019

Diploma of Software Engineering | Mumbai University | Mumbai, India | 2016

SUMMARY

As an accomplished data scientist with four years of experience, I excel in natural language processing, computer vision, time series forecasting, predictive modeling, and statistical analysis. My proficiency in Python, SQL, and big data technologies enables me to implement data-driven solutions that drive business success. I have a proven track record of delivering innovative solutions that have resulted in a substantial 30% reduction in operational costs.

EXPERIENCE

Machine learning engineer | Digitap.AI - Freelance | India | October 2022 - January 2023

- Developed a credit risk scorecard utilizing credit bureau data, achieving an 87% accuracy rate, resulting in a 15% reduction in financial risk costs. Collaborated with stakeholders throughout the project and presented key findings.
- Contributed significantly to the organization's risk management system and overall success by successfully implementing an Alternate Scoring Model using advanced techniques. The model accurately identified fraudulent customers with a 92% accuracy rate and assigned them a risk score between 100 and 900.

Senior Data Scientist | Eclerx - Morgan Stanley | India | July 2021 - September 2022

- By employing LayoutLM and Pytesseract, automated 96% of the lending team's manual tasks, streamlining loan document processing. Developed and deployed a Flask API using Jenkins and Treadmill, resulting in enhanced efficiency and reduced processing time for end-users.
- Collaborated with the team to deploy and configure Label Studio on the server, obtaining model feedback and generating automated labeling for invoice documents. This reduced the human effort required for document labeling by 80% and facilitated performance monitoring of the deployed model in production.
- Built a model for automated invoice processing using Tesseract OCR, YOLOv5, and BERT QA, significantly reducing the business team's workload by 90%. The model extracted crucial parameters such as date, taxes, and amount, providing valuable insights into spending and lending patterns.

Data Scientist | HDFC Life | India | July 2019 - July 2021

- Attained a 95% accuracy rate on the Life toolkit by training a machine learning model to extract significant parameters and verify documents such as Death Certificates, Pan Cards, and Aadhar Cards. Led a team in developing the model, collaborated with vendors, and managed annotations.
- Developed and implemented the Lie Detector model using Open Face features and LSTM, resulting in a 91% accuracy rate. This model aided the Medical Underwriter team in examining the Video MER of the customer to identify truthfulness. Conducted an experiment to collect the dataset.
- Created a bot that reduced the team's manual labor by 90%. This bot executed complex SQL queries and employed natural language processing techniques to separate data for distribution to various teams.
- Designed and developed a Qlik Sense dashboard for loan processing, which provided an intuitive and easy-to-understand visualization of the risk associated with each customer. Enabled the lending team to quickly and confidently assess customer creditworthiness.

SKILLS

Programming languages / Databases: Python, Java, SQL, Oracle, NoSQL

Libraries and frameworks: TensorFlow, Keras, PyTorch, spaCy, NLTK, Scikit-learn, Seaborn, Pandas, OpenCV, LLM, BERT, Layoutlm, LSTM, Transfer Learning, Vision Transformers, GAN, Regression, Decision trees, Random forests, SVM, XGBoost, AdaBoost, LightGBM, Clustering, PCA, t-SNE, ARIMA

BI and Analytics tools: Qlik Sense, Google Analytics, Tableau

Big Data technologies: Hadoop, Spark

Cloud platforms: AWS (EC2, Lambda, S3, Sagemaker), Google Cloud, Azure

Other tools and technologies: ETL, Kubernetes, MLflow, GitHub, Django, Flask, Jenkins, Jira, Microsoft Office (Excel, PowerPoint), Linux