

Pankaj Sharma

6503 Bluff Spring Rd, Austin, TX-78744

Email: luvnuries@gmail.com

Tel: (919) 532 6153

Data Science Engineer

Experience

As a Data Science Engineer, I believe in the power of telling and finding stories through data. I have 10 years of experience in design and development of machine learning models, data exploration, enrichment, mining, statistical analysis, building data warehouse and visualization of data for Business Intelligence solutions using different machine learning algorithms and Hadoop ecosystem - HDFS, Hive, Sqoop, Impala, Spark, Python. My current venture is leveraging Google Cloud Platform for deploying ML solution at scale using TensorFlow, Cloud ML engine, BigQuery, BigTable, PubSub, DataProc and Google Cloud Storage.

Aug'2018- Present Machine Learning Consultant, Deloitte

- ❖ Opioid Care solution on Google Cloud platform, Ingested data from different sources in cloud storage, data exploration in datalab and built warehouse to provide predictive analytic on mortality rate. Segmented counties at higher risk and recommended factor to reduce the risk.
- ❖ Developed a regression model in TensorFlow to maximize Oil production by predicting choke valve setting based on input from IOT sensors.

2013-2018 Data Scientist, Cognizant Technology Solutions

- ❖ Developed data models, algorithms to implement machine learning solutions as a member/mentor with team of data scientist, data engineer and visualization engineer.
- ❖ Ingestion of data from different source system via Sqoooping/FTPS and perform data enrichment, exploration and enrichment.
- ❖ Involved in creating/designing Hive tables, and loading and analyzing data using hive queries
- ❖ Developed **GBM** classification model for DNA mapping of more than 475 labels with 95% accuracy.
- ❖ Chatbot for Document Search Engine using **NLP, Tf-idf** and web scrapping for keyword searching using cosine similarity in urls and webpage content.
- ❖ Applied PCA algorithm to decrease the 2000 dimensions dataset to 120 Dimension with only 0.3% decrease in variance.
- ❖ Exploratory Analysis and feature engineering to best fit the model using **Python**.
- ❖ Find best model and parameters using SK-learn **Grid Search CV**.
- ❖ Initiated and Implemented deep-learning works with **LSTM** for sentiment analysis.
- ❖ **Multivariate analysis** to build an algorithm for forecasting of next quarter variance using time series analysis and **RNN** with an accuracy of 86.2%.
- ❖ Created framework to build and deploy Jars on cluster built using **SBT assembly** for different Machine learning modules such as Training, Prediction, Transformation.
- ❖ Implementation of POCs/POVs with Spark Data Frames and SQL, and Spark Machine Learning (by using objects like Transformer and Estimator for Pipelines, Evaluator, Cross Validator).
- ❖ Responsible to deliver data driven end-to-end solution, i.e. ingesting raw data, transformation, model training and generating prediction using Cloudera big data ecosystem (**Sqoop, Spark, Hue, Hive, Impala**).
- ❖ Expertise in **SK-learn, Tensor flow, Keras** and **H2O libraries**.
- ❖ Implemented data analytics projects and dashboard development on **Tableau and Qlik Sense**.

2012 ETL Architect, Tata Consultancy Services

- ❖ Extensively worked on Informatica, PL/SQL, Linux shell scripting as an ETL architect.
- ❖ Develop various Oracle **PLSQL package, procedure and Unix shell script** as per business requirement.
- ❖ Created the Data Warehouse Data Model to store atomic level data and the summarized data.
- ❖ Performed SQL performance tuning & optimized queries using Explain plan.
- ❖ Designed complex mapping architecture for critical applications with data enrichment, mapping design and flow of data.
- ❖ Architectural design, development & production support of financial applications for Back office IT operation.
- ❖ Worked on to implement the partitioning and sub partitioning methodologies to store and process high volume data in optimized fashion.
- ❖ Responsible in gathering requirements from users and designing Use cases, Technical Design and Implementation of end solutions for customers.

•

2009-2011 Software Engineer, Birlasoft

- ❖ Developed mapping to extract, transform and load data using Informatica Power center.
- ❖ RPD, report and dashboard development in OBIEE.

Education and Certifications

- ❖ **Master of Technology**, Software Engineering, *Birla Institute of Technology*
- ❖ **Google Cloud Certified Professional Data Engineer**
- ❖ **Google Cloud Certified Professional Architect**
- ❖ **A-Z Machine Learning with R and Python**, *Udemy*
- ❖ **Machine Learning** by Andrew Ng, *Coursera*
- ❖ **Natural Language Processing** (Stanford)

Technology

Python
Spark
Big Data Hadoop
Machine Learning Algorithms
NLP, Deep learning
Linux Shell scripting
SQL

Tools

Tableau/Qlik Sense - Exploration & Visualization
Intellij - Modules developments for Scala/Python
Jupyter - Data exploration and model selection in Python
Informatica Power center - Data enrichment.
Control M - Scheduler
Tortoise SVN/GitHub – Versioning
GCP – Cloud platform

References

LinkedIn: <https://www.linkedin.com/in/pankaj-sharma-99aa1921/>
GitHub: <https://github.com/luvneries>
Stack Overflow: <https://stackoverflow.com/users/8485796/pankaj-sharma>
Hacker rank: <https://www.hackerrank.com/luvneries>