

Premanand Naik

(+91) 8983118219

(+91) 7972550240

Key competencies and strengths

- Associate consultant with a curious, analytical mind having 3 plus years of experience in data science, machine learning & deep learning along with strong understanding of machine learning algorithms and data engineering capabilities.
- Experience in full life cycle of software project development in various areas like design, applications development on large scale development efforts leveraging industry standard using ETL & big data technologies like Hadoop, Hive, Spark, Kafka, Talend.
- Experience in building Machine learning & deep learning applications to solve various business problems at industry levels and presenting them to diverse audiences.
- Good understanding of deep learning architectures like RNN, LSTM, BERT, Seq2Seq modelling etc.
- Experience in solving various forecasting business problems using statistical models like ARIMA, SARIMAX and deep learning models like LSTM.
- Good understanding & hands on experience on NLP pipelines & modelling architectures.
- Hands on experience in using python and R languages for building machine learning models and data analysis.
- Hands on experience on Hadoop technology stack (HDFS, Map - Reduce, Hive, Kafka, Flume and Spark)

Education

MIT Pune, Pune University

B.Tech. Information Technology (June 2015)



Experience

Saama Technologies Inc., Pune. (Aug '19 - Present)

Associate Consultant – Big data & Data Analytics

- Designing of data pipelines to load data from Amazon S3 to Hive Tables.
- Reading S3 files using Spark, converting files from LZ to RZ and loading it into Hive tables.
- Designing of data pipelines to load data from Infomix DB to hive tables.
- Building ML models and data pipelines for business



Accenture Digital, Pune. (Oct '15 – June '19)

Senior Software Engineer – Machine Learning & Big data

- Developed GBM models in R and Python to predict iETA for shipment delivery to end customer.
- Fulfilled all Machine learning & big data engineer responsibilities for generating insights for a high-end logistics, insurance industries.
- Developed and presented iETA prediction system for Logistics giant. Achieved 10% better accuracy and reduced variance by 40% versus historical analysis.
- Designed ML/NLP powered chat bot for Insurance and Logistic giant and presented it as Co-Innovation asset to industry clients. Achieved 1st rank in chatbot Co-Innovation challenge.
- Designed NLP pipelines using deep learning architectures like sequence to sequence modelling, RNN, LSTM, BERT to understand the contextual relations between words and predicting the next word sequences.
- Developed forecasting models using ARIMA and LSTM architectures to help business forecasting the future demands.
- Collaborative participation and team leading role in Digital Co-Innovation platforms to solve real world business problems and generating insights by harnessing the power of machine learning, AI.
- Applying an optimized machine learning and statistical modeling techniques to develop and evaluate algorithms to improve performance, quality, data management and accuracy.
- Designing of spark jobs with scala and java to compare the performance of spark with hive and sql.
- Implemented spark using scala and spark sql for faster processing of real time data feeds coming from kafka.
- Using Impala to query HDFS data for faster retrieval of data.
- Importing files from AWS S3 into RDD and performing data wrangling activities and loading it into Redshift.
- Performed optimization techniques in Hive tables to reduce query latency by partitioning and bucketing of hive tables.
- Spark jobs developed to import data from Redshift into Dataframes, perform transformations and loading it into files on S3 and Redshift tables.



- Integration of Spark with RServe to run R models on distributed spark framework for faster processing of predictive models.

Projects

Intelligent ETA Prediction (Real Time) for a Global Logistics Giant, @Accenture,

- Trained Gradient boosting machine to predict estimated time of arrival of shipments with 80% accuracy.
- Trained K-means clustering to build country clusters for entire globe and achieved great reduction in variance and over fitting of data.
- Tuned Gradient boosting machine which enhanced the prediction accuracy by approx. 5 to 10 times of Random forest and decision tress.
- Optimized Python code to generate 10 different country level models.
- Statistical languages – R, Python
- Big data technologies - Spark, Hadoop, kafka, spring boot, SQL, Java, Scala, AWS Redshift, AWS S3

Chat bot for a Global Logistics Giant, @ Accenture,

- Designed voice & text featured chat bot for automating routine tasks of technical support team. Achieved strategic cost cutting of resources by 2-3 times by deploying ML and NLP based production ready chat bot.
- Increased efficiency of technical tasks with chat bot by 30 times versus performing these tasks manually by support engineers. Strategic reduction in number of client incident tickets by 10 percent after chat bot deployment.
- Customer centric UI, machine learning and NLP capabilities of chatbot has made easier for non-technical user to access business reports or other logistics related information.
- Facilitated customer with provision to expedite the shipment delivery to arrive early, if it is delayed.
- Trained Decision tress to predict the delayed shipments with 90 percent accuracy and proposed alternative MOT's (Mode of transport) and LSP's (Carrier) to expedite the shipment.

Neural machine translation – A technique powering Google translate, @ Accenture,

- Designed neural machine translation system using sequence to sequence modelling in keras to translate German sentences to English sentences.
- Trained RNN based seq2seq model over 32000 samples and validated over 8000 samples with an improved validation loss of 1.422 and average accuracy of 0.7 percent

Forecasts for product demand, @ Kaggle,

- Developed ARIMA based statistical model to forecast the demand of products across 4 warehouses for next 3 months.
- Tuned ARIMA model by Using Auto ARIMA to pick best values of model parameters and trained ARIMA model with an RMSE value of 0.16

ARIMA & LSTM time series models for S&P 500 data, @ Kaggle,

- Designed ARIMA & LSTM models to forecast the closing price and compared the performance of ARIMA and LSTM models based on RMSE metric.
- Applied gaussian filtering approach to smooth the dataset which resulted in improved performance of models. The Gaussian filtered predictions reduced the RMSE by close to 100%
- Average RMSEs for Gaussian filtered data, ARIMA: 10.98 & LSTM: 12.22

RFM analysis for customer segmentation, @ Accenture.

- Applied k-means clustering to segment customers based on RFM score and identifying potential customers of the company.
- Performed k-means clustering with 4 iterations, avg time taken 2.0035 secs, avg silhouette width of 0.33.

HP Logistics, @Accenture

- Technologies: Spark, Hadoop, kafka, spring boot, SQL, Java, Scala, AWS Redshift, AWS S3, R, machine learning.
- Designed spark job using scala api to load json files from FTP server into RDD and pushing the json record to kafka topic.
- Designed spark jobs to load streaming data from kafka topic into RDD and applied some transformations to achieve desired business outcomes and storing it back into Redshift to further layers of processing.
- Designed spring boot based spark application to implement spark sql queries on distributed

multithreaded environment to achieve faster processing of records.

- *Implemented spark connection with FTP server to receive files into RDD and loading into Redshift using spark to redshift library.*

Esure Insurance Analytics, @Accenture

- *Technologies: Talend Enterprise Data Integration, Alteryx, Putty, TAC, AWS Redshift, AWS S3, R, predictive modelling.*
- *This was migration project to migrate insurance data from Alteryx to Redshift using Talend DI.*
- *Created Talend ETL mini batch jobs to receive CSV files from AWS S3 buckets then performing some data cleaning activities and loading it into AWS Redshift tables.*
- *Implemented File Transfer Protocol operations using Talend Studio to transfer files in between network folders.*
- *Experienced in fixing errors by using debug mode of Talend.*
- *Schedule the Talend jobs with Talend Admin Console, setting up best practices and migration strategy.*
- *Created Context Variables and Groups to run Talend jobs against different environments like Dev, Test and prod.*
- *Design and developed end-to-end ETL process from various source systems to Staging area, from staging to Data Marts.*
- *Experienced in integrating java code inside Talend studio by using tJavaRow, tJavaRow, tJavaFlex and Routines.*
- *Experienced in writing SQL Queries and used Joins to access data from MySQL, Hive & Redhsift.*

Motiva Data Lake, @Accenture

- *Technologies: Talend Big Data, Putty, TAC, Hive, Impala, Hadoop, HDFS, MySql*
- *Created Talend ETL mini batch jobs to receive CSV files from FTP server then performing some data cleaning activities and pushing it into HDFS.*
- *Designed Talend ETL jobs to get files from HDFS, load it into staging tables in Hive.*
- *Designing of ETL's to perform transformations on source Hive tables depending upon the business requirements and loading the achieved results into curated layer tables for reporting purposes.*
- *Implemented File Transfer Protocol operations using Talend Studio to transfer files in between network folders.*
- *Experienced in fixing errors by using debug mode of Talend.*
- *Schedule the Talend jobs with Talend Admin Console, setting up best practices and migration strategy.*
- *Created Context Variables and Groups to run Talend jobs against different environments like Dev, Test and prod.*
- *Design and developed end-to-end ETL process from various source systems to Staging area, from staging to Data Marts.*

Accomplishments

Accenture Co-innovation Contest Winner, 2018, First winner in Accenture's innovation contest for designing AI based chatbot.

ACE award winner @ Accenture, 2018, Accenture excellence award winner.

International Karate Gold Medalist, 2018, Gold medal winner in Kata and Kumite.

Skills

Programming Languages *Java, Python, R programming, Scala*

Machine Learning Packages *Tensorflow, Scikit learn, Keras*

Distributed Computing *AWS, Hadoop, Hive, Apache Spark*

Machine learning algorithms *linear regression, logistic regression, k means clustering, and Gradient Boost machine*

Deep learning algorithms *RNN, CNN, LSTM, Seq2Seq*

Forecasting models *ARIMA, SARIMAX*

ETL tools *Talend Big data, Talend DI*