**Premanand Naik**

premanand.naik09@gmail.com

+91 8983118219, +91 7972550240

[www.linkedin.com/in/premanand-naik-83866377](http://www.linkedin.com/in/premanand-naik-83866377)

**Experience**

**Consultant @ Capgemini**, Pune, India, 12/2019 – present

My responsibilities in the data science team were:

* Developing the machine learning models to extract paragraph from unstructured text and cluster them based on different control engineering loops.
* Designing of information retrieval system using Spacy NLP, MITIE that resulted in new contextual patterns and relations between control blocks in control engineering technical design documents.

**Associate Consultant @ Saama Technologies**, Pune, India, 07/2019 – 12/2019

* Liaising with the clients and stakeholders to understand their requirements, challenges etc. and keeping them up to date on the progress of proposed solution.
* Designing of data pipelines using spark jobs to extract data from oracle, performing transformations and loading them into hive tables.
* Optimization of spark jobs and use of parallelization to increase the efficiency of data transformations and data loading by almost 30 percent faster compared to legacy systems.

**Senior Software Engineer @ Accenture,** Pune, 10/2015 – 06/2019

* Providing high-end consulting to Supply chain business clients to help them sharpen their business strategy by implementing machine learning and statistical models.
* 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.
* Designing of data engineering pipelines using big data technologies like Hadoop, Spark, Hive etc. and consulting clients across supply chain, insurance & oil industries by providing data driven solutions using applied machine learning and data science solutions.
* Developed machine learning model using GBM in R and Python to predict iETA for shipment delivery to end customer with an accuracy of 82 percent and resulted in improved on time shipment deliveries by almost 50 percent.
* 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 and also improved performance by using different activation functions and advanced optimizers. Also applied different NLP techniques like tf-idf, word2vec
* Designing data driven solutions using data science and analytics by leveraging big data technologies and machine learning, and proposing them to leadership by visualization and storytelling techniques.
* Performed predictive analytics and designed various ML applications using Stochastic models, Bayesian Modeling, Classification Models (SVM/Random Forest/Decision Tree), Cluster Analysis, Neural Network, Non-parametric Methods, Multivariate Statistics along with their performance optimization and analysis.

**Education**

**MIT Pune, University Of Pune**, Pune, Maharashtra, India

***Bachelor of Engineering, Computers & IT Major***, June 2015

Grade: distinction

**Courses and Trainings**

**Machine Learning**, by Stanford, @Coursera

**Neural Networks and Deep Learning**, by Deeplearning.ai, @Coursera

**Convolution Neural Network**, by Deeplearning.ai, @Coursera

**Deep learning specialization**, by Deeplearning.ai, @Coursera

**Technical Skills**

* Experience in solving various forecasting business problems using statistical learning models like ARIMA, SARIMAX and deep learning models like LSTM, neural networks.
* Proficiency in building applications using tools (R/Python) and use of appropriate Python/R libraries (e.g. pandas, numpy, Keras, TensorFlow, nltk, matplotlib, sklearn (scikit-learn), ggplot2 etc.) and delivering business values to clients.
* Experience in supervised and unsupervised machine learning algorithms such as liner regression, logistic regression, k-NN, decision tress, random forest, PCA, GBM, XGBOOST, k-means, SVM, bagging and boosting techniques. Well versed with NLP techniques like tf-idf, bag of words, word2vec, glove embedding’s.
* Strong experience in working with distributed data systems like Hadoop and related technologies like Spark, Hive, PySpark, kafka, SparkR
* Experience in building recommender systems for supply chain management. Hands on experience in deep learning & state of art NN architectures like CNN (ConvNet), RNN, BERT, LSTM, neural networks.

Tools:

* Programming Languages: **Java, Python, R programming, Scala**
* Machine Learning Packages: **TensorFlow, sklearn, Keras, pytorch**
* Distributed Computing: **AWS, Hadoop, Hive, Apache Spark, Amazon Redshift**
* Python Libraries: **pandas, scikit-learn, numpy, nltk, spacy, matplotlib, seaborn, flask**
* R libraries: **Caret, caTools, Rshiny, ggplot2, dplyr**
* Machine learning algorithms: **linear regression, logistic regression, k-means, Gradient Boost machine / gbm, xgboost, decision trees, random forest**
* Deep learning algorithms: **RNN, CNN, LSTM, ConvNet, BERT, neural networks, State of art**
* Forecasting models: **ARIMA, SARIMA**
* Cloud based analytics platform: **azure, GCP, AWS, AWS Sagemaker, Google Cloud**
* Others: **pyspark, sql, HQL, nosql, mysql, natural language processing, openCV, Tesseract-OCR, Analytical Skills, Client Relations,, Builds Relationships, Problem Solving, Demand Forecasting, Interpersonal Skill, information retrieval, data mining, Consulting**
* Operating system: **Windows 10, UNIX, Linux**

**Functional Skills**

* Experience in Insurance/ Automotive/ Supply chain/ Engineering services/ Digital industry domains
* Hands on experience in analytics delivery, through the entire life cycle of an analytics engagement
* Client facing experience along with requirement understanding, UAT and enhancements
* Onsite team engagement experience
* Engaged in client interactions, participating in technical workshops
* Developing models, mentoring junior team members, building models as per standards, creating intermediate documents and getting the work approved with stakeholders as per plan

**Awards**

**ACCENTURE Co-Innovation Contest,** 2018

First winner in Accenture’s innovation contest for designing AI based chatbot.

**ACCENTURE Excellence Awards,** 2018

Accenture excellence award winner.

**International Karate Championship,** 2017

Gold medal winner in Kata and Kumite.

**Projects**

**Control Narrative system for Schneider Electric,** 2020

Designing of information retrieval system using Spacy NLP that resulted in findings of new patterns and relations in technical design documents. Knowledge extraction based on contextual relation understanding using NLP techniques and word embedding’s. Creation of pseudo code from knowledge extract using parse tree and instructions. Keyword extraction of equipment and IO tags from design documents and creation of visualization using PySimpleGUI python library.

**Intelligent ETA Prediction (Real Time) for a Global Logistics Giant,** 2019

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 & R code to generate 10 different country level models.

**Chat bot for a Global Logistics Giant,** 2018

Designed voice & text featured chat bot for automating routine tasks of technical support team using reinforcement learning. 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.

**HP Logistics,** 2017

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.

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

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

**Digit Recognizer, Image recognition using CNN**,

Developed CNN based deep learning model to detect the digits out of images with an improved accuracy of 0.98 with Kaggle rank of top 42%.

**Sentimental analysis – IMDB movie reviews using BERT NLP,**

Predicting the sentiments of IMDB movie reviews as Positive, Negative, Neutral using BERT pre-trained model with tensor-flow as backend. Achieved overall accuracy and AUC score 0.86.

**Crypto currencies price prediction,**

Developed deep learning model to predict Bitcoin and Ethereum market prices using LSTM model. Trained LSTM model of 20 neurons, 8 features with 50 epochs and Linear as activation function. Achieved MAE of 0.04 with LSTM model.