Vipin Indoria

Gurugram, Haryana, India

 [vipinindoria@gmail.com](mailto:vipinindoria@gmail.com) +91-9166670029  [linkedin.com/in/vipin-indoria](https://www.linkedin.com/in/vipin-indoria)

# Summary

* Having 4.6 years of experience in design and implementation of highly scalable, efficient, robust & fault-tolerant data science and engineering solutions from the ground up.
* Experience in customer's business value and technical landscape discovery, solution architecting, costing, infra sizing, project management and team handling.
* Experience in creating regression models, feature engineering pipelines, predictive models and using data mining algorithms to deliver insights for building action oriented solutions to the complex business problems.
* Experience in building reusable and metadata driven components for data integration using REST APIs and message queues, data processing pipelines using DataOps approach with TDD for batch and real time data using lambda and kappa architectures.

# Experience

## Technical Lead - Big Data Engineering & Data Science

### Team Computers

Nov 2018 - Present (2 years 4 months +)

* 1. Transport Analytics Centre: Conceptualized and Built TAC using Google Cloud Platform for the logistics department of a cement manufacturing customer in order to monitor the driver’s driving behaviour, driver risk categorization and driver’s training program.

Tech Stack: Debezium, Pub-Sub, GKE, Storage Bucket, Dataflow, BigQuery, Apache Beam, Python, Cloud Composer, SVM, Scikit-Learn, Machine Learning, Classification

* 1. Telematics Command Centre: Built Data Platform to monitor agri/construction machines in real-time to get insights about machine health, utilization and service ticket using Azure Cloud Platform.

Tech Stack: Debezium, Kafka, ADLS, Databricks, Python, SQL, NLTK, Natural Language Processing

* 1. Insurance Data Lake: Drove a big data project to conceptualize and built insurance data lake. Tech Stack: Hadoop, Hudi, Kafka, Spark, Ranger, Kerberos, Postgres, Scala, Airflow, Debezium
  2. Smart Factory Accelarator: Managed QSFA product development which is designed to initiate a result- oriented approach that facilitates prevention of any possible collapse in the machinery.

Tech Stack: TensorFlow, Keras, LSTM, Python, Docker, Kafka, Spark, Postgres and Qlik

## Software Engineer - Big Data & Machine Learning

### Confluxsys

Feb 2018 - Nov 2018 (10 months)

1. Fraud Analytics Framework: Built a fraud detection framework to detect the anomalies in user-access relationship inside an IAM system.

Tech Stack: Auto-ML, Clustering, Java, Oracle, JSAT, Apache Spark

1. Text Analysis Framework: Conceptualized and built a framework to detect duplicate roles created in an identity and access management system having the same list of entitlements but with different name. Also created an interface to check grammatically correct sentences.

Tech Stack: Jaccard similarity matrix,Oracle, Java, Stanford NLP, Natural Language Processing

## Software Engineer - Big Data & Data Science

### EATLAB

May 2017 - Feb 2018 (10 months)

1. Built a data platform to integrate various data sources (camera feed, touchscreen table web portal, load cell IoT data) into a unified storage sink and process the data for feature engineering and descriptive analytics for understanding the consumer’s eating behaviour.

Tech Stack: TensorFlow, Python, Darknet, Pub-Sub, BigQuery, Storage Bucket, CNN

1. Built a framework to forecast the future sale of food/drinks and to analyse the effect of advertising. Tech Stack: Python, Scikit-learn, Tensorflow, Keras, Statsmodel, Regression, Time-Series Forecasting

## Associate Project Engineer - Data Science & Data Engineering

### CDAC R&D

Sep 2016 - Apr 2017 (8 months)

1. Involved in the development of big data and deep learning sandbox for supercomputing clusters. Worked on a bioinformatics project to analysing the genetic variations (SNPs – Single Nucleotide Polymorphisms) across different varieties of rice genomes with origin from different countries using big data and machine learning frameworks.

Tech Stack: Clustering, Variant Spark, Python, Tableau

# Education

## Centre for Development of Advanced Computing

### Postgraduate Diploma, Big Data Analytics

2016 - 2016

## B K Birla Institute of Engineering & Technology, Pilani

### B.Tech. (Honours), Electrical and Electronics Engineering

2011 - 2015

# Licenses & Certifications

 **Data Engineering** - Qwiklabs

[https://www.qwiklabs.com/public\_profiles/b8ee5c85-dc74-4b3e-a7c7-b499bb130960](http://www.qwiklabs.com/public_profiles/b8ee5c85-dc74-4b3e-a7c7-b499bb130960)

 **Advanced ML: ML Infrastructure** - Qwiklabs

[https://www.qwiklabs.com/public\_profiles/b8ee5c85-dc74-4b3e-a7c7-b499bb130960](http://www.qwiklabs.com/public_profiles/b8ee5c85-dc74-4b3e-a7c7-b499bb130960)

# Skills

Python (Programming Language) • Apache Spark • Hadoop • Scala • Google Cloud Platform (GCP) • Microsoft Azure • Machine Learning • Deep Learning • Java • SQL