



# VIPIN SIRIAH

Data Science | MBA | Statistics

Data Science Practitioner with 7+ years of experience in IT business services. Expertise in machine learning, statistical analysis, deep learning, and data modeling with 5+ years of hands-on in Python & SQL; exposure to R, Google Analytics, Tableau, SPSS & SAS.

✉ siriah.vipin@gmail.com

📞 08698853718

🌐 [linkedin.com/in/vipin-siriah-177a9834](https://www.linkedin.com/in/vipin-siriah-177a9834)

🐙 [github.com/vipinsiriah-git](https://github.com/vipinsiriah-git)

## PROJECTS

### Dynamic Discounting using Reinforcement Learning

#### MakeMyTrip

05/2021 - Present

Bangalore, India

Travel Super-App

Flights Dynamic Discounting

- Employed reinforcement learning models that optimize discounts for certain coupons so as to maximize the expected revenue for each search case while maintaining market share.
- Created models to predict the conversion probability of the user booking a flight based on the discount provided.
- Exposed the ML model and the RL agent using Falcon API that can handle up to 100 thousand post-requests per hour while maintaining response time within 100 ms, as per SLA.
- Created data pipelines using Apache Airflow-controlled DAGs to extract, transform and load the data science abstraction tables using the high amount of clickstream data getting logged by various teams across the flights business.

### Infrastructure Recommendations

#### Xoriant Solutions

12/2020 - 04/2021

Bangalore, India

Client : Morgan Stanley Capital International

Achievements/Tasks

- For a leading investment bank, developed data engineering pipeline to extract and transform the Splunk data from SQL into Hive.
- Designed and developed ML algorithms to predict personas of the new users coming into the system based on user features like job-description, resume text etc.
- Developed ML algorithm to identify the users which need an upgrade in hardware and primary machine based on the system and application performance data.

### Automated Media Monitoring: Content Classification

#### Wipro Digital

05/2020 - 11/2020

Bengaluru, India

Audio and Text Classification

Achievements/Tasks

- Created a data-pipeline to create and store the live streaming data from the TV-Eyes portal which includes XML parsing, Audio Clipping, and Transcript Extraction. Used Python + MySQL pipeline for implementing this data ingestion.
- Created DB structures like Triggers, Stored Procedures, Materialized Views to enable data flow and to identify the duplicate contents. Created UI to get the labels against the media contents.
- Extracted Audio features from the radio streaming data using the librosa library and text features from transcript data.
- Trained a 3-Layer ANN on Audio MFCC & Spectrogram features of audio and an XGBoost Classifier on Sentence Vectors of the transcripts to classify between Ad and Non-Ad content.

## ORGANIZATIONS

### MakeMyTrip (05/2021 - Present)

Lead Data Scientist

### Xoriant Solutions (12/2020 - 04/2021)

Senior Data Scientist

### Wipro Digital (09/2019 - 11/2020) [🔗](#)

Lead Data Scientist

### Tata Consultancy Services (06/2014 - 09/2019)

Business Analyst - Data Science

## SKILLS

python

machine learning

deep learning

reinforcement learning

athena

airflow

redshift

data modelling

statistics

falcon api

nlp

mysql

rdbms

finance

business intelligence

aws

hadoop

anomaly detection

dataiku

google analytics

hypothesis testing

cognitive-search

hive

pyspark

explainable AI

retail analytics

## TOOLS & PACKAGES

### Machine Learning

pandas, numpy, scikit-learn, statsmodel, keras, tensorflow, scipy, SHAP, LIME, ELIS, iblearn, xgboost

### NLP

spacy, nltk, gensims, stanford-NLP, GloVe, Word2Vec, elastic-search, solr, sklearn-crfsuite, pdfminer, lxml

### Data Engineering

IBM DB2, Oracle RDBMS, SQL Server, MySQL, MongoDB (NoSQL)

### Business Intelligence/ Visualisations

tableau, powerBI, excel, plotly, bokeh-plot, matplotlib, seaborn, streamlit

## EDUCATION

### Post Graduate Program in Management

Great Lakes Institute of Management, Chennai [🔗](#)

05/2017 - 04/2018

3.14/4.00

Finance & Analytics

- Full-Time Residential Program with Majors in Business Analytics and Finance.

- Selected based on 93 percentile score in CAT 2016-17.

## PROJECTS

### Revenue Guidance using Deep Learning

#### Wipro Digital

12/2019 - 06/2020

Bengaluru, India

Revenue Guidance for a CPG Major

- Lead a team to design time-series prediction algorithms using an ensemble of LSTMs to predict revenues for multiple brands across geographies for a major American clothing company. [↗](#)
- Designed and developed the data model for data requirements of the Revenue Prediction Enterprise Solution using Oracle DB, optimizing the efficiency of the existing Database System.
- Employed Explainable AI techniques to understand and explain the deep learning model behaviour and analyse the causality of model outputs.

### Incentive Optimisation Algorithm

#### Wipro Digital

09/2019 - 01/2020

Pune, India

Insurance Renewal Prediction

- Created statistical models to predict the renewal of general insurance policy for insurance domain clients to replace the SAS based legacy client system.
- Used these predictions to calculate the incentives to be given to each insurance agent to maximize the revenue by applying operational research algorithms provided by the Scipy package.
- The aim is to improve the legacy Incentive Optimization process and streamline the process to reduce the turnover time by integrating the algorithm to the workflow web-portal using Flask API.

### Lead Scoring Algorithm ( POC)

#### Wipro Digital [↗](#)

09/2020 - 01/2020

Bengaluru, India

Client Domain: Telecommunication - Eastern Europe

- Responsible for providing technical solutions for the business problem, creating and delivering technical presentations internally and externally.
- Developed Lead Scoring model and it's database architecture for the clients in the telecommunication domain.
- The model scores the leads, based on their probability of conversion into a sale using the pre-sales lead data in the CRM database.

### TCS Decision Fabric - Connected Clinical Trials

#### Tata Consultancy Services [↗](#)

03/2018 - Present

Pune, India

Product that automates the Clinical Trial Processes

Achievements/Tasks

- Contributed to the development and enhancement of TCS AI-based product Decision Fabric DF (see link above), which provides a broad spectrum of solutions to solve cognitive automation-related problems in the pharmacovigilance domain, using Python for development, and MongoDB in the back-end. [↗](#)
- Responsible for presenting the solution to multiple Pharmaceutical clients, delivering upgrades for the client-specific requirements, and designing the enhancements to integrate it with client systems.
- Developed Connected Clinical Trials (CCT) (see link), a component of TCS DF, which provides a solution to extract structured data from unstructured PDF files using text-mining and Named Entity Recognition techniques (NER).
- The tool extracts information in standard E2B R3 and IDMP formats from unstructured reports found in a variety of formats.
- Employed NER techniques using bi-directional LSTMs, used packages like Keras, TensorFlow, Spacy, etc to achieve desired results.

## EDUCATION

### PG Diploma in Applied Statistics

#### Indira Gandhi National Open University

02/2020 - Present

Courses

- Mathematics and Statistics (Currently Pursuing)

### Bachelor of Engineering

#### GHRCE, Nagpur University, India

07/2010 - 05/2014

8.50/10

Courses

- Full-time Bachelor's Degree
- Majors in Electronic & Telecommunication
- Selected based on AIEEE 2010 score

## ACHIEVEMENTS

### Skilienza - WhattheHack Hackathon (03/2020) [↗](#)

Appeared amongst the top-10 in the hackathon problem by skilienza (see link above)

### NMIMS SAS Datathon (11/2017)

Finalist at Business Analytics and Data Mining championship- 2017 held at NMIMS by SAS Global.

### OpenSoft at IIT-KGP (01/2012)

Finalist at national level software competition at 'Open-Soft' held at IIT-KGP in Kshiti-2012.

### Competitive Exams (2014 - 2016)

Cleared GATE 2014 with score 468 and 95 percentile and scored 93 percentile in CAT 2016-17.

## CERTIFICATES

1. Neural Networks and Deep Learning by Andrew N.G. - Coursera [↗](#)

2. Improving Deep Neural Networks: Hyper-parameter tuning, Regularisation and Optimisation [↗](#)

3. Structuring Machine Learning Projects [↗](#)

4. Sequence Models and NLP using deep-learning by Andrew N.G.

5. R Programming by Johns Hopkins University [↗](#)

6. Sentiment Analysis with Deep Learning using BERT [↗](#)

## LANGUAGES

English



Hindi



## INTERESTS

Trekking

Hackathons

Cricket

Biking

Cycling

History

## PROJECTS

### Text Summarization using Lexical Chaining

#### Tata Consultancy Services - TRDDC

09/2016 - 03/2017

##### Achievements/Tasks

- Developed text summarization algorithms using concepts of similarity-based Lexical Chaining and Text-Rank in Python.
- Worked to rewrite the code of summarization algorithm in python which was published in C# originally as a part of research by the researchers at TRDDC.

### Anomaly Detection (Unsupervised)

#### Tata Consultancy Services

08/2015 - 02/2016

Pune, India

##### Client Domain : Banking and Financial Services

- Developed unsupervised anomaly detection algorithms using features extracted from the SWIFT messages used for electronic fund transfers.
- Used Isolation Forest algorithm to identify anomalies, the algorithm provides an isolation score for each transaction, based on the features extracted from SWIFT messages.

### NLP based Free Text Search

#### Tata Consultancy Services

03/2015 - 06/2015

##### Achievements/Tasks

- Developed an Ontology-driven search that converts user free text into existing business entities, thereby creating SQL search query for user text query.
- The algorithm identifies entities, attributes values, and condition on them from the free-text query and translates it into a structured SQL query.

### Cognitive Search Algorithm for Service Desk

#### Tata Consultancy Services

Client Domain: Hi-Tech Service Platform - US

##### Achievements/Tasks

- Developed a document recommendation system for clients in the Hi-Tech domain to predicts relevant documents for a given query and textual inputs of a user.
- The top-10 documents are recommended to the service desk agent for solving their tickets.
- Employed an algorithm based on multi-class classification for getting probabilistic scores and elastic-search similarity-based scores for each user query.
- The prediction algorithm is used as an API by the Service Platform Web-portal.

### Full Stack Web-Development

#### Tata Consultancy Services

06/2014 - 02/2015

Client Domain: Automotive Industry -USA

##### Issue Tracking System Website

- Responsible for designing and maintaining ASP.NET based website using enterprise-level design methodologies in the ASP.NET framework.
- Used ASP.NET- MVC framework to enhance the existing web application and JavaScripts for front-end development.
- Handled database design and development for the web-application using MS SQL Server and IBM DB2.
- Worked with web technologies and relational databases as a developer, gained expertise in SQL.
- Recognized by the clients for optimizing performances of multiple database objects in SQL DB2 to improve overall efficiency and performance of the Issue Tracking website.

## VOLUNTEER EXPERIENCE

### Karma-Yoga (Student Initiative)

#### Great Lakes Institute of Manament

05/2017 - 03/2018

Chennai

##### Tasks/Achievements

- As part of the Karma Yoga team at Great Lakes Institute of Management, Chennai actively participated in CSR activities like education drive, plantation drive, help creating entrepreneurship opportunities for rural communities near the Campus

### Volunteer

#### Robin hood Army

11/2019 - Present

Bnagalore

##### Tasks/Achievements

- Responsible for teaching the students on weekend and helping them in there schools curriculum