

# VIMAL JOSE

Email: vimaljosehere@gmail.com | Cell: +61 449 757 507 | Melbourne | Australia PR (189) | Notice Period: 30 days

## LEAD DATA SCIENTIST (Role: Analytics Lead for Asia)

Data Science (Python & R) | Cloud (AWS) | Tableau | Statistics | SQL | Big Data | Data Warehouse Modelling | Agile

Seasoned analytics professional with 9 years of experience in Data Science, Python, R, Tableau, SQL, AWS Cloud, Data modelling, Big data and agile. Winner of the ACE, Best IC and Platinum awards and consistently rated in the top 5%.

### CERTIFICATIONS

- ① AWS Certified Solutions Architect Associate
- ② EMC Data Science Proven Professional (Expert)
- ③ Certified Tableau Associate
- ④ Certified Scrum Master
- ⑤ Deep Learning L1 (IBM CC)

### TECHNICAL SKILLS

- ✓ Data Science in Python & R
- ✓ Tableau Desktop 10.3
- ✓ Hypothesis and A/B Testing
- ✓ Excellent SQL/PLSQL skills
- ✓ Exploratory Data Analysis
- ✓ Productionizing ML models
- ✓ Data Cleansing & Wrangling
- ✓ Feature Engineering
- ✓ Spark and Hive Hands-on
- ✓ Deep Learning (TensorFlow)
- ✓ PostgreSQL and NoSQL DBs
- ✓ Cloud Solution Architecture

### EDUCATION DETAILS

MBA in Finance	2020 – 2022	Deakin University	Ongoing
MS in Data Science (Research)	2018 – 2020	Liverpool John Moore Uni	CGPA 3.1/4
PG Diploma in Data Analytics	2017 – 2018	IIIT	CGPA 3.9/4
B. Engg in Computer Science	2007 – 2011	Mahatma Gandhi University	76.59%

### PROFESSIONAL SUMMARY

- Currently performing the role of **Asia Analytics Lead** with teams across Philippines, Malaysia and India.
- Responsible for **building data science teams from scratch** and **architecting their infrastructure** and processes.
- **Successfully led** and contributed to 10+ end-to-end analytics projects across all phases of SDLC/CRISPDM
- **Excellent mathematical and intuitive understanding of statistics and machine learning algorithms.**
- Actively worked with the end user in capturing and **converting user requirements to data science use cases.**
- Good knowledge of classification, Regression and time series models including **Linear and Logistic regressions, Decision Trees, SVMs, Kernel SVMs** and **ARIMA** and **SARIMAX** models.
- Successfully implemented **RandomForest, XGBoost** and other **ensemble models** in Python and R
- Proficient in creating **interactive visualizations**, dashboards and scorecards using **Tableau (Certified)**
- Experienced in developing **complex SQL queries** and PL/SQL packages for data replication.
- Excellent skills in **data cleansing**, imputation and **data wrangling** using **tidyverse(R)** packages.
- Have successfully scalably **productionized Neural Network models** in Tensor flow using **AWS**.
- Proficient in **Ensemble** techniques, **model stacking** and **cross-validation** implementations.
- Have worked on **PyTorch** and **TensorFlow** models using **Convolutional Neural Networks** for Image classification
- Worked on **Generalized Sequential Pattern (GSP) models** and **apriori models** for IVR optimization.

---

## PROFESSIONAL HISTORY

---

***DXC Technology.***

*Asia Analytics Lead, Data Scientist 3, 2019–Current*

---

### ***Projects details***

- |  |  |  |
|--|--|--|
| 1  | <b>Project:</b> Operations Engineering Data Science Team | <b>Environment:</b> Python, SQL, AWS, Power BI |
| <ul style="list-style-type: none"><li>• Successfully scaled up the team from 3 to 12 members within 8 months in the capacity of Asia Analytics Lead.</li><li>• Setup the Analytics team <b>processes and documentation structure</b> and created an <b>initial pipeline</b> by gathering and <b>prioritizing requirements</b> from multiple stakeholders.</li><li>• <b>Developed data catalogs and dictionaries</b> for multiple disparate data sources and setup a data discovery process across teams so that data scientists and analysts can easily plan for multi-source data analyses.</li><li>• Implemented a <b>predictive capacity planning model</b> for stable accounts based on a time series daily call volume prediction with <b>SARIMAX and two exogenous Fourier terms</b>, with an accuracy of 74%.</li><li>• Developing a new set of <b>KPIs and metrics</b> for the analysis of clients' incident data to provide the account directors an overall view and <b>building predictive analytics</b> on the metrics for <b>optimization and benchmarking</b>.</li><li>• Was a key contributor in <b>building the Operations engineering data science team</b> and <b>architecting</b> the Data Science team's <b>cloud infrastructure on AWS</b>.</li></ul> |  |  |

***Flipkart Internet (Acquired by Walmart at \$20B).***

*Lead Analyst, 2018–2019*

---

### ***Projects details***

- |  |   |  |
|--|---|--|
| 1  | <b>Project:</b> Incident Journey Scoring System               | <b>Environment:</b> Python, R, SQL, Hive, Tableau        |
| <ul style="list-style-type: none"><li>• Was <b>key in the project ideation</b> and formulation by brainstorming with key stake holders - VP for Customer Experience, Director of Analytics and respective team managers.</li><li>• Conducted the <b>Exploratory Data Analysis</b> and <b>Inferential Statistics</b> on the data set including escalations, incidents, order and user data.</li><li>• Developed the project with two analysts and was <b>responsible for planning and conducting the pilots and A/B tests</b> before deploying the model to production.</li><li>• Built a <b>customer anxiety scoring model</b> based on incidents, orders and users data for implementing a predictive escalation and incident prediction system which has resulted in a 11.8% reduction in escalations and a 6.3% reduction in incidents. <b>RandomForest</b> gave the best results out of all the models tried, at an accuracy of 74.8%.</li></ul> |   |  |
| 2  | <b>Project:</b> Customer Experience Metrics Reporting (L0/L1) | <b>Environment:</b> Python, R, SQL, Hive, Tableau, Excel |
| <ul style="list-style-type: none"><li>• Was responsible for the correct reporting, prediction and monitoring of all the metrics of the CX team.</li><li>• <b>Presents and co-ordinates</b> the weekly and monthly meetings for the customer experience team with the leadership which includes the CFO (only for monthly) and Senior Directors.</li><li>• <b>Was awarded "Rock star – Value addition" by the business leadership</b> for the improved rigor in the monthly metrics and the innovative projects that I drove, including resolution rate prediction for a future view of the business metric and maturity models for metrics with long-maturity times.</li><li>• Was responsible for capturing and converting the visions of the business leadership into projects and analyses for the analytics team.</li></ul>  |   |  |

***Dell EMC Corporation (EMC<sup>2</sup>).***

*Consultant, 2016 – 2018*

---

### ***Projects details***

---

1	<b>Client:</b> Qatar based Telecom major <b>Environment:</b> Python, R, SQL, SAP Business Objects, Tableau, Greenplum <ul style="list-style-type: none"> <li>Was part of requirements phase itself - Travelled to client place, met the stake holders including Country Heads and successfully converted user requirements to clear data science and analytics use cases.</li> <li>Built clustering and classification models to determine the upsell and cross sell opportunities in mobile packs using <b>Kernel SVM</b> (RBF kernel gave the best accuracy, we tried <b>Random Forest</b> and <b>XGBoost</b> as well)</li> <li>Completed a thorough <b>Exploratory Data Analysis</b> and <b>Inferential Statistics</b> for the CDR data for use in models.</li> <li>Successfully led the requirement gathering, analysis, development and testing of <b>BI reports in Tableau and BO</b></li> <li>Designed and developed the semantic layers for analytics and <b>Data warehouse Modelling</b>.</li> <li>Built a <b>collaborative filtering model</b> for a recommendation engine for personalized marketing which improved the purchase rates by 11% and with an intent accuracy of more than 40%, over a span of 3 months.</li> <li>Used <b>A/B testing</b> to check the efficiency of one of the UX designs over the existing design.</li> <li>Built a proof of concept on <b>Restricted Boltzman Machines</b> for recommendation engines.</li> <li><b>Led the team of 8</b> in developing <b>PostGRESQL functions in green plum</b> for ETL (Massively Parallel Processing DB)</li> </ul>
2	<b>Client:</b> Leading Bank in APJ <b>Environment:</b> Python, R, SQL, Tableau, Spark, Hive <ul style="list-style-type: none"> <li>Successfully built <b>Generalized Sequential Pattern (GSP)</b> and <b>Apriori</b> models for IVR optimization to achieve ~30% faster response times using a backend of Hive External tables.</li> <li>Architected and designed the <b>data pipelines for Data Science</b> as well as BI reporting.</li> <li>Designed the dimensional model for reporting as well as created Tableau and SAP BO reports and dashboards for performance analysis at multiple granularities.</li> <li>Successfully did <b>data cleansing</b> and <b>wrangling</b> on IVR data using <b>tidyverse</b> packages including <b>outlier identification and imputations</b> for Customer Call data.</li> </ul>
3	<b>Internal:</b> Machine Learning Bootcamp <b>Environment:</b> R <ul style="list-style-type: none"> <li><b>Was 1 of the 3 trainers who conducted Machine Learning Bootcamp for the whole Data Engineering Division</b></li> <li>We covered topics from basics of R programming to algorithms like <b>Naïve Bayes</b>, <b>K-Means Clustering</b>, <b>KNN Classification</b>, <b>Dimensionality reduction techniques like PCA, LDA and t-SNE</b> and time series analysis techniques like <b>Holt-Winters</b> and <b>ARIMA variants</b> and advanced techniques like <b>DBSCAN</b> and <b>Expectation Maximization (Gaussian)</b> clustering methods.</li> <li>We also discussed statistical concepts like <b>Correlation analysis</b>, <b>ANOVA</b>, <b>pair-wise t-tests</b>, <b>z-scores</b>, <b>confidence intervals</b>, <b>skewed data</b>, <b>unbalanced classes</b> and <b>multicollinearity</b>.</li> <li><b>This was very well received and we received appreciation from APJ heads for the effort.</b></li> </ul>
4	<b>R&amp;D:</b> Security Analytics <b>Environment:</b> Python, R, Tableau <ul style="list-style-type: none"> <li>Modelled ARIMA models to detect anomalous behavior in networks by analyzing domain controller log data Used <b>AdaBoost</b>, <b>Random Forests</b> and <b>Model stacking</b> for correctly classifying user behavior after aggregating over a weekly time span.</li> <li>Used <b>statistical tests</b> to estimate the significance of the readings from two different domain controllers.</li> </ul>
5	<b>Client:</b> EMC Product Team (VX-Rail) <b>Environment:</b> TensorFlow, Python, R, Tableau <ul style="list-style-type: none"> <li>Forecasted the internal metrics using <b>incremental learning</b> using <b>ADAM</b> optimizer with <b>TensorFlow</b> (Tried <b>Stochastic Gradient</b>, <b>Nesterov Accelerated Gradient (NAG)</b> also)</li> <li>Used <b>multi-threaded queue runners</b> to achieve accelerated data loading for <b>out-of-the-core learning</b>.</li> <li>Improved the speed of predictions by               <ul style="list-style-type: none"> <li><b>L1 regularization</b> for generating a <b>sparse matrix</b> (Tried <b>ElasticNet</b> also with multiple r values also)</li> <li>Downsizing the float variables to 16 bit.</li> <li>By tuning the <b>dropout</b> parameter so that the neurons that sustain generalizes well.</li> </ul> </li> <li><b>Also took part in scalably productionizing this model using REST APIs and redis for Message Queueing.</b></li> </ul>
6	<b>R&amp;D:</b> Video Analytics <b>Environment:</b> Python, OpenCV

- Participated in the R&D project on constructing an analytics framework on videos.
- Successfully implemented pedestrian recognition with ~70% accuracy using **HOG descriptors and Linear SVM** on CCTV feeds.
- Generated the heat map based on human activity levels from CCTV feeds, so as the customer could do the placing of the store better.
- Implemented a Proof Of Concept for a scalable facial recognition architecture with **HOG Descriptors** and **128 bit face embeddings** using pillow and openface in Python.

**Infosys Ltd.**

*Associate Consultant, 2014 – 2016*

**Projects details**

- |   |   |   |
|---|---|---|
| 7 | <b>Client:</b> British Energy Major   | <b>Environment:</b> R, BO 4.1, Tableau, SQL Server 2014, Oracle 11G |
|   | <ul style="list-style-type: none"> <li>• Worked with the <b>advanced analytics team</b>, and completed <b>data cleansing and Exploratory Data Analysis</b> for more than 7 assignments.</li> <li>• Gained expertise in <b>Regression models (Linear and Logistic), Residual Analysis, Regularization techniques and Bias-Variance trade-off.</b></li> <li>• Successfully developed <b>classification models with Naïve Bayes and Random Forests with F-Scores &gt; 0.75</b></li> <li>• Was part of the end-to-end project planning and worked through all phases of SDLC for reports and universes.</li> <li>• Was involved in the project planning, analysis, design and the development of 20+ universes and 100+ reports.</li> <li>• Working closely with the clients and highly appreciated by the client for pro-active problem solving skills.</li> <li>• <b>Leading a team of 6 members and conducted trainings to cross train SQL developers into BO.</b></li> <li>• Working closely with Tableau Team, Database team and the BO team to identify where to place business rules.</li> <li>• Successfully delivered a project which had to be completed in ~65% of the estimates due to an unforeseen circumstance by leading the team in agile mode and by detailed delegation of tasks.</li> <li>• Successfully introduced automation of the testing of universes by guiding a VBA developer from bench for adapting the freely available VBA code to suite our requirements.</li> <li>• 7 appreciations in a single year from clients, managers and team members and <b>rated with the top rating.</b></li> </ul> |   |

**Accenture Ltd.**

*Software Engineering Analyst, 2011 - 2014*

**Projects details**

- |   |   |  |
|---|---|--|
| 1 | <b>Client:</b> Nordic Banking Major   | <b>Environment:</b> BO 4.0, SAP Dashboards, Oracle 11G           |
|   | <ul style="list-style-type: none"> <li>• Was involved in planning, requirement analysis and leading the development of some of the reports.</li> <li>• Successfully did advanced performance tuning at universe and report levels with runtime savings of up to 400%.</li> <li>• Spearheaded the continuous improvement of the universes by investigating better than existing solutions.</li> <li>• <b>Led a team of 5 members including colleagues at the same level.</b></li> <li>• Actively worked with the end user in capturing requirements and finalizing the functional specifications.</li> <li>• Defined the test cases to check the correctness of the data in the reports by comparing them with backend data.</li> <li>• Implemented the drill down hierarchies for multi dimensional analysis as per client requirements.</li> </ul> |  |
| 2 | <b>Client:</b> German ERP company   | <b>Environment:</b> BO 4.0, SAP Dashboards, Infocubes, BEx, HANA |
|   | <ul style="list-style-type: none"> <li>• Involved working very closely with the Client Solutions Architect and <b>leading a team of 3.</b></li> <li>• Designed, developed and modified universes on top of SAP Infocubes.</li> <li>• Developed 7 WebI reports on top of BEx queries and universes built on SAP data providers.</li> </ul>   |  |

- Contributed in all the phases of Software Development Life Cycle (SDLC).
- Implemented a new process flow for the design of WebI reports which was highly appreciated by the client.
- Prepared the security architecture of the implementation including row level and group wise restrictions.
- Developed reports by linking different data providers in a single report, and made all possible incompatible objects compatible by merging the common objects.
- Implemented index awareness in the universe to avoid issues with few duplicate labels.
- Created master detail reports with dynamic hyperlinks to sub-reports.

- **Received the 'ACE' award, the highest individual award for delivery excellence at Accenture.**
- Participated in the complete Software Development Life Cycle (SDLC).
- Modified and optimized the performance of 4 universes and created 9 reports.
- Designed and maintained complex universes, resolved loops and traps by using contexts and aliases.
- Implemented aggregate awareness for improving the efficiency of the reports.
- Improved consistency and reduced code with the use of @functions (@aggregate aware, @select, @where etc).
- Modified complex PL/SQL packages for data transformation in Oracle streams replication.
- Configured Oracle Streams for ETL and tested the data for correctness and concurrency.
- Used index awareness for reducing the load time of prompts in a large volume table.
- Was involved in the performance testing for the ETL and reports before moving to production.

- Developed 8 reports and supported the development of 5 universes on top of SAP Infocubes.
- Used BIWS and QaaWS connections **to develop an innovative approach for making dashboards schedulable.**
- Contributed to Accenture Asset Library on ‘Scheduling dashboards’ which was used by 10+ projects.
- Defined aliases and contexts to resolve loops and tested them for correct data retrieval.
- Created and modified existing hierarchies in the universes to meet the drill analysis requirements.
- Worked extensively with BO functionalities such as Breaks, Sections, Ranks, Variables, Alerters, Filters and Sorts.
- Adopted strategies for LOVs involving customizing the LOVs, Cascading LOVs and exporting LOVs with universe.
- **One of the five most awarded new joiners of 2011 at Accenture, Chennai. (Performance points wise).**

- Permanent Resident of Australia – 189: Can work anywhere in Australia.

- Certified Neuro-Linguistic Programming practitioner and a passionate learner of evolutionary psychology.
- Have conducted multiple sessions on “Influencing people with Neuro-Linguistic Programming” including one session in the Dell EMC Annual Leaders meet.
- Was the trainer for ‘Machine Learning Bootcamp’ at Dell EMC, Bangalore.
- Regular speaker at Toast Masters
- Hypnotist in training.

- **Rohit Kapoor**, Victorian Government, Melbourne - +61 478 527 580

- **Murali Nagaraj**, Dell EMC, Sydney

- +61 439 787 778