

Selva Prabhakaran

Skills

Languages	R, Python
Core Competencies	R Programming, Develop data products, Machine Learning / Statistical Modeling, Problem solving, Predictive Models, Parallel computing, High performance computing, Forecasting Accuracy Enhancement, Discrete Event Simulation
packages	InformationValue on CRAN, WOE on github
Domains	Automotive, Supply Chain, Marketing Analytics, Manufacturing, Logistics, Finance
Others	Git, RMarkdown, Sweave, Wordpress, RStudio Server

Education

- 2004–2009 **M.Sc.(Hons)**, *Birla Institute of Technology and Science, Pilani*, Pilani, India.
Biological Science
- 2004–2009 **B.E.(Hons)**, *Birla Institute of Technology and Science, Pilani*, Pilani, India.
Mechanical

Experience

- 2014–current **Senior Information Analyst**, *IHS Inc*, Bangalore, India.
- Developed the critical parts of forecasting algorithm that generates monthly automotive forecast and market shares for 30 US based car manufacturers.
 - Achieved and proved > **90% accuracy for 12M out-of-sample**. when current industry standard is 75%.
 - Leveraged novel machine learning algorithms, parallel computing and efficient code to achieve the high accuracy while maintaining minimal code run-time.
 - Developed and maintained the optimised production code for the IHS Automotive subscription based forecasting product.
 - Responsible for Creation, Delivery and Updates of Production Code (in R) for Polk - IHS Automotive <U+2018>s Sales and Market Share forecast product that leverages advanced analytics, that is the industry standard for all US Car brands.
- 2011–2014 **Information Analyst - 2**, *Caterpillar Inc*, Chennai, India.
- As part of the Information Analytics Center of Excellence, worked in multiple advanced analytics projects for Cat facilities and product groups worldwide.
 - Developed the demand forecasting algorithm for prime products that is currently deployed in-production.
 - Interact with stakeholder at all levels, gather requirements in case of open-ended projects, create the problem framework and solution approaches, design and develop prototypes and deliver results and recommendations.
- 2010–2011 **Business Analyst**, *Mu Sigma Business Solution Pvt Ltd*, Bangalore, India.
- Performed predictive modeling, scoring to determine the SMB customers who are 'at-risk' of attrition from Dell.
 - Co-ordinated with clients and delivered scorecards and actionable insights for the APAC and US region on a day-to-day basis.
 - Responsible for on-time delivery, first pass quality and consistency of business sensitive deliverables.

Projects

- Github Open source work and repositories on <https://github.com/selva86>
- R Package 'InformationValue' is an R package that provides companion function for analysing the performance of classification models. Also, provides function to optimise probability cut-off score based on used specified objectives, Plot 'ROC' Curve in 'ggplot2', 'AUROC', 'IV', 'WOE' Calculation, etc to aid accuracy improvement in binary classification models. 'InformationValue' is available as a package on CRAN <https://cran.r-project.org/web/packages/InformationValue/index.html>
- Automotive Sales Forecasting Developed the core part of the forecasting algorithm for IHS Automotive's subscription based forecasting product (powered by Polk), used by marketing teams of 30+ major US based car manufacturers. **Achieved consistent out-of-sample accuracy of >90%, where industry standard is 75%.** Also, optimised the code run time from **1hr to 10min** and maintainer of the production code. #Machine learning #HighPerformanceComputing
- Prime Products Demand forecasting Forecast the 12M sales volume of large excavators (LHEX) using in-field machine utilization data and economic indicators data. Large Excavators division is a product family that has accumulated inventory due to poor sales. Forecasting the sales volumes accurately is highly critical for this product division and other related product groups, dealers and suppliers. **Achieved high out-of-sample accuracy of 90% and proved potential savings of \$19M.** #Forecasting #R
- Cat Logistics Parts Forecasting Forecast the demand of various components and parts based on economic indicators and product usage data. This was done for central logistic distribution center catering to facilities in all global regions. The accuracy of the forecast was highly business sensitive and critical for success. #R #Timeseries Regression #Forecasting
- Scheduling and Throughput Optimization Involves tooling and scheduling optimization to enhance the throughput of CNC based machining centers of 8+ Mazaks. The algorithm optimally assigns what parts and tools should be assigned to which machines, based on demand and tool requirement. As a result, the average waiting times due to tooling and fixture changeovers is minimised. In case of a machine failure, it also suggests alternate machines to load the part into. #AlgorithmicOptimisation
- Text Mining Manufacturing facilities conduct a lot of tests on before approving a part. There is a lot of test data collected from these tests, that it is difficult to manually read through them and understand the main causes of errors and failures. Text mining was used to build an automated mechanism that can quickly extract the topics, root causes and recurring problems that occur during testing. #R, #LDA, #TextMining
- Discrete Event Simulation Used discrete event simulation to model entire facilities, warehouses, supply chain networks to identify bottlenecks and optimise resources and aid capacity planning. This analysis was done for multiple Caterpillar facilities world-wide. #DES #BottleneckAnalysis #OperationsResearch
- Safety stock modeling Finished parts have a longer lead-time (time between order and delivery) and a higher cost compared to its component parts. So, enough quantity has to be stocked to cover for the consumption during the lead-time. The lead-time of the finished part is governed by the its component with longest lead-time, sometimes, the longest lead-time parts cost very less, so storing these components in bulk is a better idea than to produce them. The aim of the project is to build a DES model of a inter-related part-supplier network, run experiments for stocked quantities at various nodes, compute the total cost at each node and finally recommend safety stock at each node (supplier). #DES #OptimumLotSizing #StrategicInventoryPlanning
- Supplier Performance Analytics Score the large number of suppliers based on performance KPIs such as On-time delivery, On-Time shipping, Defects, Past Due shipments. It is used to identify poor performing suppliers so as to improve point-of-use availability and in some cases, development of alternate suppliers. #R #SupplierAnalytics #LoessRegression

Cost modeling - Statistical Estimate the expected cost of a part based on its design features using advanced statistical techniques using R. It also estimates the cost impact of making a design changes to a part. It helps in purchasing decisions by helping to identify right supplier for a part keeping in consideration the manufacturing, logistic and geographical constraints. Compared to traditional methods of cost estimation, these were added benefits and proved to be highly effective especially when making purchasing and sourcing decisions. #R #PredictiveModeling

Interests

Languages Love learning, problem solving and trying out new programming languages
Blogging I maintain an advanced Rstats website at <http://rstatistics.net/>
MOOCs Algorithms, Statistics, Data Visualization and Machine Learning.
Sports Good at and love to play Badminton, Cricket and Table tennis