**RAJU PENMATCHA, PhD, MBA**

**Data Science, Big Data, Technology**

609-456-0844, vispe73@gmail.com, US Citizen

**SUMMARY**

* Nineteen years of experience in Architecting, developing and Managing teams in Enterprise software solutions including **Machine Learning**, **Artificial Intelligence** (**AI)**, **Big Data** technologies, **Hadoop**, **Machine Learning** Supervised and Unsupervised Learning, Text Analytics and **Unstructured data** analysis. Worked on various **Recommendation Engines** and several Java/J2EE, .Net, TIBCO based platforms.
* Work experience includes assignments as Senior Big Data Executive, Data Scientist/Architect, Enterprise Architect, Mentor at Bank of America Merrill Lynch, Johnson & Johnson, HSBC, Mayo Clinic, American Express, DuPont, Goldman Sachs, Federal Reserve Bank (NY), Merrill Lynch, Bank of America, Bell South, Accenture

**SKILLS**

Languages: Python, R, Scala, Pig, Hive, Java/J2EE, JSP, C#, ASP.Net

Database: HBase, M7, HDFS, Oracle, DB2, Sybase, SQL Server, JDBC, ODBC, Toplink, Hibernate, SAP R/3

Development: Hadoop, MapReduce, Spark, Mahout, Data Modeling, Yarn, Zookeeper, J2EE, EJB, SOAP, REST

Tools Used: DataRobot, Eureqa, H2O.ai, Scikit-learn, Pandas, Numpy, Neo4J, Graph Database, Tableau, MemCache, RCurl, RDD, JDK, Tibco, Matlab, Revolution R, Flume, Sqoop, Unica, Oozie, Datameer, Platfora, MapR, Cloudera, BigInsights, Azure, AWS, EC2, EMR, Pandas, Numpy, Anaconda, Spyder, Jupyter, Mapplotlib, Seaborn, Lucene, Elastic Search, Solr, Sklearn

Data Science: Supervised Learning, Unsupervised Learning, Deep Learning, Reinforced Learning, Regression, Lasso, Ridge, KNN, KMeans, Clustering, Neural Networks, SVM, Naïve Bayes, Fuzzy Logic, NLP, NLTK, Spacy, Topic Modeling, Decision Trees, Random Forest, Bagging, Boosting, GBM, AdaBoost, TensorFlow, SkFlow, Algorithmia, CNN, Deep Learning, BigInsights, ARIMA

**EXPERIENCE**

**DataRobot, New York, NY** May 2017 – present

**Customer Facing Data Scientist**

DataRobot is an Automated Machine Learning product firm. Helped DataRobot acquire new clients working with decision makers, conducting Proof of Concepts, nurturing prospects into clients and developing collaborative culture.

* Ensured alignment with key technology and business stakeholders across globally diverse, Agile teams.
* Helped Bank of America through building predictive models in FX, Fixed Income, Investment Banking, Research, Commercial Lending, Wholesale Credit, Client Relations, M&A.
* FX Volume Prediction: built time series models to predict daily FX volume for CLS at a minute level using data from primary exchanges - EBS, Reuters, BofA volume, bid and ask rates, spreads, VWAPs, simple and exponential moving averages, order book entries, etc. Used Bollinger Bands, MACD, market events, holidays for EUR/USD, USD/CAD, etc currency pairs. MASE values were impressive compared to a naïve model.
* Customer Attrition Prediction: built a highly successful customer attrition predictive model with 80% accuracy on FICC electronic trading from Bloomberg terminals using time series, feature engineering with financial ratios, etc.
* Capital Review Committee Revenue Prediction: predicted yearly revenues for years 1 to 3 for the bank on 16 products ranging from Treasury, Advisory, Credit to FX, and had beat bankers estimates.
* Funded Loan Growth Prediction: developed predictive models for funded loan growth for the Corporate Banking group at Bank of America and improved prediction probability six times. Found key drivers and early indicators.
* Worked with Balyasny Asset Management (BAM) hedge fund, JPMC, TD Bank, GRA (Global Risk) at BAML.
* Built custom machine learning models on large datasets in use cases such as optimal capital allocation, commercial loan growth, customer attrition, market trend prediction for the bank.
* Built multi-class sentiment analysis models on bank’s research reports using NLP and Spacy.
* Moved Machine Learning projects into production and created tangible value for the firms.
* Brought business insights showing feature interactions in ratings tables, prediction explanations.
* Built several workflows that combined data preprocessing steps with feature engineering, feature selection, model selection, hyper-parameter tuning, model stacking, blending, using cross validation to avoid overfitting, validating models with lift charts and ROC curves, explaining insights through feature importance analysis, partial dependency plots. Handled class imbalance and large datasets. Explored human – machine hybrid approaches.
* Captured trends, seasonality patterns through time series models such as ARIMA, used lag variables and sliding window techniques, feature engineered variables through iterations.
* Analyzed unstructured text in analyst reports, built sentiment analysis using TFIDF, NLP, Spacy.
* Balanced algorithm accuracy over speed in XGBoost, Random Forest, GLM, ENet Blender, Logistic.
* Worked with bank regulators on variable stress testing, parameter sensitivity analysis.
* Built challenger models for BAML regulators, the Model Review Management group, a three month long process, with variable stress testing, hyper-parameter sensitivity analysis, out of time validation, and model deployment.
* Helped Humana insurance with Marketing mix optimization, Emergency Room attendance estimates.
* Developed Oil recovery models for Hunt Oil, and transport ETA predictions for Rail Inc. and BASF.
* Handled large scale transactional, trading, loan, hospital, transportation, oil production data.
* Evangelized Artificial Intelligence, Machine Learning through presentations, online webinars, blogs.
* Wrote popular blogs on Machine Learning and received company’s special award on content creation.

**Environment:** Python, R, SKLearn, Time Series, ARIMA, Multiclass, Anomaly Detection, Feature Engineering, Imbalanced data, SQL, Hive, Hadoop, Tableau, Spacy, NLP, Spark, DataRobot, Eureqa, Nutonian

**SRP Systems, Princeton, NJ** July 2015 – May 2017

**VP - Data Science & Big Data**

SRP is a data analytics solution company. **Clients: HSBC, StorQuest, Zoetis, Mayo Clinic.**

Lot of hands-on works and Managed technical team as **Data Scientist** and Big Data Architect on various client projects listed below. Our **machine learning algorithm** that rates healthcare providers **won a third price** from **Robert Wood Johnson Foundation** contest, New Jersey.

**Anti-Money Laundering (AML) Data Science, HSBC, Wayne, PA**

* Developed real-time, transaction screening SaaS software that reduce false positives and false negatives and detects risk before a financial crime happens. Intelligent, data science driven SaaS software on the cloud or on-prem for transaction/customer screening, insider threat, FCPA.
* Driven by machine learning, natural language processing, text analytics, link analysis using graph databases.
* Linked hidden PII data such as email ids, phone numbers, addresses, friends, relatives, work relations, etc.
* Used public source data, sanctions list, OFAC lists, entity resolved private vendor, data from KYC processes, from correspondent banks. Used 3 Vs: volume, value, velocity of transactions on SWIFT, Chip type transactions.
* Did holistic risk assessment and risk intelligence. Built prescriptive systems where systems aid decision making.
* Highly flexible and configurable platform with easy integration capabilities into Hadoop based data lakes as well as advanced engines based on data science and AI.
* Developed predictive models using cluster analysis, anomaly detection, and feature selection through PCA for dimensionality reduction using their large transactional and KYC datasets.
* Helped bring external data from vendors such as Factset, NexisLexis.

**Dynamic Pricing in Retail using Data Science, StorQuest, Los Angeles, CA**

* Helped this retail-chain stay competitive through cutting edge pricing analytics models from large scale data.
* Developed demand prediction models using supervised machine learning models such Random Forest Regressor, Extra Trees Regressor, SVM, XGBoost, GBM, Logistic Regression.
* Worked on large sets of sales, transaction, marketing, and pricing data. Conducted data analysis/cleaning, predictive analytics, handled inventory sold outs, generated price elasticity curves.
* Applied **price optimization** using advanced price optimization libraries in Python and **dynamic pricing** models for revenue management

**Text Analytics and NLP, Zoetis, London, UK**

* Stimulated their sales using Natural Language Processing (NLP) and creative fuzzy logic-based text analytics to build MDM solution matching the Pharma company and its vendors.
* Used Neo4J (Graph Database) for speed and reinforced learning techniques in AI to build a highly advanced machine learning based MDM (Master Data Management) solution.

**Big Data Architecture and Data Security, Mayo Clinic, Rochester, MN**

* Secured Mayo Clinic private data using data masking and Big Data architectures for areas such as **genomics**, **radiology**, and **clinical reports** using **Machine Learning** and **predictive analytics**.
* Put together an architecture in place for them using Azure cluster with HDP, data security using Kerberos, Atlas, Ranger, Knox (perimeter security), Data Masking, KMS

**Environment:** Latent Dirchlet Allocation (LDA), Topic Modeling, Semantic Analysis, NLP, TF/IDF, Elasticsearch, AWS Cloud, Python, R, Pipeline, Grid Search, Azure, Supervised Learning, Optimization, Pyopt, Matplotlib, SQL, Pandas, Numpy, SVM, RandomForest, ExtraTrees, Boosting, Bagging, Neural Networks, Deep Learning, Cloudera, Spark, MLLib, PySpark, Big Data Security, Elastic Search, Lucene, H20

**American Express, New York, NY** 2014 – 2015

**Senior Big Data Scientist and Architect**

* Worked ‘Recommender’ type projects that used Big Data technologies and collaborative filtering techniques such as cosine similarity to make real-time offers and merchant recommendations to card members. Descriptive, predictive, and prescriptive analytics were developed.
* Collaborated with several lines of business to develop a state of the art analytical modeling platform that allowed modelers to collaborate on development of machine learning models such as KNN, GBM, SVM, Logistic Regression, Random Forest, Clustering, Recommendation Systems on a large Hadoop infrastructure with Governance and Security
* Drove development of a customer centric Expert Feedback System (**Customer 360**). Pulled together data from all custom contact points of Amex with respect to campaigns, customer response, fulfillments into Big Data analytics.
* **Evangelized** Big Data technologies and their capabilities to multiple lines of business at American Express and there by influenced their adoption and value creation in creating personalized offers to customers, better recommendations to merchants, and data driven marketing campaigns.
* Worked on Recommendation Engine type projects that used collaborative filtering techniques like cosine similarity to make real-time offers and merchant recommendations to card members.
* Worked on **Digital Partnerships** with TripAdvisor, Yelp, etc. Developed MVPs.

**Environment:** Hadoop, HBase, NoSQL, Hive, Pig, Map Reduce, Yarn, H20, Spark ML, RDD, Scala, Mahout, Recommender, KNN, Bayes, Linux, MapR, Innovation, Revolution R, Python, Big Data, Solr, Lucene, Netezza, MPP, Shell Scripts, RCurl, Amazon EMR/S3, MemCache, RDD, Platfora, Tableau, IBM BigInsights, Scikit Learn

**Axalta (DuPont), Glen Mills, PA**  2013 – 2014

***Senior Enterprise Architect / Big Data & Cloud***

Leadership Responsibilities:

* Put together strategic and technical direction in transitioning DuPont systems and applications into Axalta domain, a **$15 million program** with **60+ people** working across the globe.
* Architected implementation of **Tibco** as middleware to integrate Magento (E-Commerce) system to SAP in the backend, and SFDC, YADA, JMS, Liferay in the frontend
* Adeptly managed conducting POCs on Tibco adapters to old SAP systems such as SAP 4.0b and 4.6c on BAPI calls and **iDocs** received from **SAP**, and in configuring SAP transports
* Put together Architectural artifacts (**TOGAF**) with redundant, fault tolerant systems with disaster recovery. Implemented load balance through F5, with a proxy server in DMZ, defined firewall rules, with Web Service HTTP and HTTPS requests coming from Salesforce (SFDC), YADA, JMS sitting in the cloud
* 300,000 end users will use these applications once completed, in three continents and in multiple languages

**Environment:** **Tibco**, Tibco BW, EAI, Axway, Magento, Liferay, SAP, Tibco SAP Adapter, Oracle, SQL Server, EDI, Gentran, Axway, Hyperion, Informatica, Microsoft SCOM, Hyper-V, Cloud, Server Virtualization, Agile, Zohodocs, ForgeRock, SAML, WebAgent, ITIL, TOGAF, Zachman, **Hadoop,** Hive, MapReduce, HBase, NoSQL, J2EE, PHP, MS Project, HortonWorks, Windows, Linux, **Cloudera**

**SmartWorks, Edison, NJ** 2005 – 2012

**CTO**

SmartWorks is a technology solutions company. **Clients included: Merrill Lynch, Federal Reserve Bank NY**. Brought 1000% growth to the company, propelled it into an INC 5000 listed firm, followed by a profitable merger. Led 100+ dotted line reports.

* Conducted vendor selection and management. Motivated tech teams through thought leadership, career development programs, hackathons and through hiring highly qualified individuals and appropriate incentives.
* Developed standards technology using client server, highly configurable, loosely coupled systems, using service oriented architectures.
* Extreme programming and agile development processes were used doing sprints in a scrum process.

Some significant projects are listed below:

1. **Merrill Lynch, Hopewell, NJ MDM Project**

Worked as an Enterprise Architect to assist them with developing SDD (Software Design Document) to manage integrate huge amounts of data between Merrill Lynch and Bank of America systems using Master Data Management (MDM) concept.

* Architected systems that support customer centric tools using BPM tools such as TIBCO, with a .Net frontend and TIBCO/Java backend with the use of SOA and ESB.

2. **Federal Reserve Bank, New York, NY ITS**

* Development using Struts, JSP, Javascript on the front end, Stateless session beans for transaction management and security, XA driver for distributed transactions, Oracle and DB2 and Mainframe in the back end. Design patterns like Session Façade, DAO, DTO, Singleton, Factory, Strategy, Decorator, etc were used.

**Environment:** J2EE, EJB, JSP, Servlets, Struts, WebSphere 6.1, 5.1, RAD 6.0, WSAD 5.1, DB2 6.0, DB2 Client, Mainframe, Oracle 9i, SWIFT, EDI formats, JUnit, Log4j, JavaScript, Design Patterns,Toad, CVS, WinCvs, AJAX, DOJO, Axis, XML, Rational Rose, iText, Ant, Use Cases, Hibernate, Spring, C#, ASP.NET 2.0, SQL Server 2005.

**Goldman Sachs, New York, NY** 2003 – 2005

**Senior Data and Software Architect**

Worked on four projects at Goldman Sachs (Private Wealth and Asset Management Group) in the roles of Architect, Project Manager, and Implementer. Year-end evaluation received: Far Exceeded Expectations.

* Project Manager handling task assignment, progress check, and management of onshore consultants (4) and offshore team (6).

**Environment:** Java, J2EE, JSP, Servlets, EJB, JBoss 3.02, WebLogic 6.0, Eclipse 3.0, CVS, NetBeans, iText, Sybase

**iSoft Group, Dallas, TX** 2000 – 2003

**Modeling and Agile Lead**

iSoft is a technology services company. **Clients included: Accenture, Bell South, Nokia, Southwest Airlines**. Worked as an Architect, Mentor and Senior Developer on multiple client projects as listed below:

**1. Federal Reserve Bank, New York, NY**

**Lead Developer; International Direct Deposit**

Did design work that involved coupling Struts with their in-house design. Modified their design to allow easy functional tests just below the Struts layer. Used design patterns like Command, Business Delegate, Visitor, Template, DAO, DAC, Value Objects and Anti-Patterns like Concern Slush, Tier Leakage.

**Environment:** J2EE, EJB 1.1, WebSphere, Oracle, JUnit 4.0, JTest, Log4j, JavaScript, Design Patterns, Anti Patterns

**2. Bell South/Accenture, Birmingham, AL**

**Architect/Mentor/Sr. Developer**

Followed all 12 practices of Extreme Programming (XP) including test-first and pair-programming

**Environment:** J2EE, Java, JSP, Servlets, GRNDS framework, Struts, Taglibs, Jbuilder 7.0, Oracle 9i, Toplink

**Mobil Oil, Dallas, TX**  1998 – 2000

**Analytical Modeling Senior Researcher**

* Developed analytical mathematical models to predict well productivity at any inclination and in 3D space.

**TECHNICAL PAPERS & PUBLICATIONS**

* Webinar: [Enhancing Customer 360 Models with Automated Machine Learning (2018)](https://www.datarobot.com/webinar/enhancing-customer-360-models-automated-machine-learning/)
* Blog: [An Engineer’s View: How to Pick an Automated Machine Learning Product](https://blog.datarobot.com/an-engineers-view-how-to-pick-an-automated-machine-learning-product) (2018)
* Blog: [Act now on Artificial Intelligence](https://www.linkedin.com/pulse/act-now-artificial-intelligence-raju-penmatcha-phd-mba/) (2017)
* Blog: [An Easy Way to Get your Company Started on Data Science and Machine Learning](https://blog.datarobot.com/easy-way-get-company-started-data-science-machine-learning) (2017)
* Blog: [Microsoft Azure HDInsight for your Big Data needs](https://www.linkedin.com/pulse/microsoft-azure-hdinsight-your-big-data-solution-penmatcha-phd-mba/) (2016)
* Blog: [Analytics at Home - A Crazy Thought, Right?](https://www.linkedin.com/pulse/analytics-home-dont-miss-raju-penmatcha-phd-mba/) (2015)
* My experience with Extreme Programming (Agile) – A Case Study (2005)
* Effects of Grid Systems on Predicting Well Productivity SPE Journal (2000)
* Use of Analytical solutions to improve Simulator accuracy, SPE Journal (1999)

**EDUCATION**

1. MBA (Marketing, Entrepreneurship), Wharton Business School, Philadelphia 2013
2. Ph.D. (Engg, Minor: CS), Stanford University, Stanford, CA (GPA: 3.9/4.0) 1998

* TOGAF 9 Foundation certified Enterprise Architect
* PMP certified from PMI