**Abhi Pramanik**

**Data Scientist**

**Dell EMC certified Data Scientist (EMCDS), Advanced Analytics Specialist**

[www.linkedin.com/in/abhidata](http://www.linkedin.com/in/abhidata), <https://github.com/pramabhi>

[abhi.pramanik1@gmail.com](mailto:abhi.pramanik1@gmail.com)

**(609) 375-8812**

**Professional Summary:**

* A Data Scientist with **9 years** of experience in analytics in highly quantitative environments in **banking, telecom, retail and insurance industries.**
* Thorough grounding in all phases of data analysis (both **structured and unstructured**), including definition and analysis of questions with respect to available data and resources, overview of data and assessment of data quality, selection of appropriate predictive models, statistical tests and presentation of results.
* Extensive programming skills in analytical programming languages such as **R, Python.**
* Expertise in machine learning using **R-packages, Scikit-learn, SPSS, Spark ML, Spark MLlib** etc.
* Well versed in machine learning algorithms such as **Linear and Logistic Regression, Random Forest, Naive Bayes,**

**Singular Value Decomposition, K-Means, K-nearest neighbors, Neural Networks** etc.

* Ability in using **Apache Spark (PySpark)** for data processing (SparkSQL) and machine learning (ML, MLlib) with big data
* Experience in implementing data analytics solutions in **AWS (EC2, EMR. S3), Google Cloud**.
* Experience in Deep Learning on GPUs using python libraries e.g. **TensorFlow, Theano, Keras** etc.
* Understanding of different components of **Hadoop ecosystem** such as Hue, Pig, Hive, HBase, HDFS, Map-reduce,

Flume etc.

* Proficient in **Natural Language Processing (NLP)**, **Text Analytics** etc. using R and Python (**NLTK, genism, TextBlob** etc.)
* Expertise in **Social Network Analysis** **(SNA)** using python (**NetworkX**).
* Expert in Visualization and dashboards using **Tableau, R Shiny, ggplot2, matplotlib and seaborn.**
* Experience writing advanced **SQL programs** for joining multiple tables, sorting data, creating SQL views, creating indexes and metadata analysis. Experiences in performing Ad Hoc queries for various application and reports on a daily, weekly and monthly basis using complex SQL queries.
* Sound knowledge in Enterprise **Data Warehousing and Business Intelligence**, **ETL (Extract, Transform, Load),**

dimensional/Hierarchical data modeling, Data mapping, Data Dictionaries [Used IBM **Cognos BI** tool].

* Strong Communication and Presentation Skills substantiated in past assignments with developers, project managers, subject-matter experts, stakeholders, system implementers, and application end-users.
* Good problem solving, reporting and statistical skills with excellent communication and inter-personal skills. Demonstrated ability to work both in independent and team-oriented environments.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Operating Systems** | Windows 10/7/XP, Linux, Unix |
| **Databases** | Oracle, SQL Server, HDFS, Microsoft Access, DB2, MySQL, Cassandra |
| **Statistical** | Microsoft Excel, R, Python |
| **Business Intelligence Tools** | * IBM Cognos,Tableau |
| **Machine learning** | Classification, regression, clustering, anomaly detection, feature engineering |
| **Business Tools** | MS Visio, MS Project, MS Access, MS Office Suite |
| **Languages** | SQL, PL/SQL, Python, R, UML |
| **Natural Language Processing** | Stanford CoreNLP, NLTK, Gensim, Apache Spark, SciKit-learn |
| **Methodologies** | Agile, Scrum, Waterfall, Lean Six Sigma |

**Certifications:**

* **Certified Data Scientist (EMCDS), Advanced Analytics Specialist from Dell EMC Corporation**

[**See Certificate**](https://www.certmetrics.com/dellemc/public/verification.aspx?code=EYPY1L14L9LPFL56)

* **Certified Data Scientist Associate (EMCDSA) from Dell EMC Corporation**

[**See Certificate**](https://www.youracclaim.com/badges/5b7029f8-f5be-4e81-8be5-60fe5f6d835f)

* **Certified Oracle Database SQL Expert from Oracle Corporation**

[**See Certificate**](https://www.youracclaim.com/badges/5675e126-effa-45c2-b458-e84c6e4a37ca/linked_in_profile)

**Professional Experience:**

**Client: Verizon, Ashburn, VA June 2018 – Till date**

**Role: Data Scientist**

**Description:** Verizon Wireless, is an American telecommunications company which offers wireless products and services. It is a wholly owned subsidiary of Verizon Communications. The project is in the pricing department. The work includes data extraction and analysis for B2B business of Verizon. Forecasting revenue, building data pipeline are some of the major areas of work.

**Responsibilities:**

* Translated business challenges into math/statistical hypotheses and prepare analysis plans to highlight potential value for the business
* Developed mechanism to run model in the past to compare and evaluate effectiveness of any changes performed in existing model building algorithm using pyspark.
* Research and application of machine learning & natural language processing for text analytics for live fraud detection on distributed systems supporting massive rates of data transfer
* Mined scanned pdfs to extract, process and load data in structured format.
* Used natural language processing to analyze and find intent in texts extracted from
* Managed and analyzed large data sets using statistical tools and techniques.
* Worked closely with the data engineering team to integrate the Machine Learning solution in production environment
* Worked on big data using PySpark and performed machine learning using ML, MLlib packages
* Trained the machine learning/deep learning models in Google Cloud Platform
* Implemented a LSTM Neural Network for forecasting revenue using keras/Tensorflow

**Environment:** Oracle, Cassandra, Linux, SQL, Hadoop, SPSS, Spark, Keras, TensorFlow, AWS, Pig, Hive, Windows 10, MS Excel, Tableau, VBA, R, Python, Google Cloud, Github, HP Quality Center, HP Unified Functional Testing.

**Client: RealNetworks, New York, NY April 2017 – May 2018**

**Role: Data Scientist, Lead**

**Description:** RealNetworks, Inc. is a provider of internet streaming media delivery software and services based in Seattle, Washington, United States. The company also provides subscription based online entertainment services and mobile entertainment and messaging services. The project is in new product development in the mobile messaging division. The product provides a platform to categorize text messages in real-time using advanced big data analytics.

**Responsibilities:**

* Research and application of machine learning & natural language processing for text analytics for live fraud detection on distributed systems supporting massive rates of data transfer
* Create and apply machine learning algorithms and techniques on a large scale of text data
* Advise and help Business Intelligence team for reporting & data visualization using Tableau
* Built machine learning models to categorize SMS traffic off-line as well as real-time.
* Conducted training and proof of concept (POC) work
* Managed and analyzed large data sets using statistical tools and techniques.
* Worked closely with the data engineering team to integrate the Machine Learning solution in production environment
* Worked on big data using PySpark and performed machine learning using ML, MLlib packages
* Trained the machine learning/deep learning models in EC2 & EMR and saved them in S3 of AWS
* Implemented a Convolutional Neural Network for a multiclass text classification problem using keras/Tensorflow
* Performed network analysis on mobile messaging traffic data to derive patterns and other useful insights

**Environment:** MySQL, Cassandra, Netezza, Linux, SQL, Hadoop, SPSS, Spark, Keras, TensorFlow, AWS, Pig, Hive, Windows 7, MS Excel, Tableau, VBA, R, Shiny, Python, Google Cloud, IBM Jazz, JIRA.

**Client: Chubb group of Insurance Companies, Warren Township, NJ Nov 15 – Mar 2017**

**Role: Data Scientist**

**Description:** Chubb Limited formerly Chubb Corp. is the largest property and casualty insurer in the United States, with over 120 offices located in 54 countries, and offers commercial, property, casualty, specialty and personal insurance services. The project was on developing, executing, tracking and analyzing targeted marketing campaigns. Utilized the social media campaign management application to develop and report on complex, multi-step campaigns. Analyze campaign performance, report on key business metrics and develop insights through customer analysis.

**Responsibilities:**

* Carried out segmentation, building predictive models and integrating secondary and primary data using R, python, SPSS and SQL
* Built classification models that predict the probability of a customer’s response to a cross-sell campaign using python.
* Documented methodology, data reports and model results and communicated with the Project Team / Manager and other data scientists to share the knowledge on retention analytics
* Took End-to-end ownership of designing, developing and deploying machine learning models (data preparation => variable selection => model building and evaluation => deployment)
* Analyzed large data sets for reporting and visualization using R.
* Tracked campaigns and communicate campaign performance and ROI analysis.
* Responsible for managing monitoring and coordinating claims fraud risk management projects
* Performed Sentiment Analysis using social media and survey data to address customer grievances and brand awareness using Natural Language Processing (NLP).
* Analyzed and manipulated high volume of the transactional data in cloud (AWS)

**Environment:** Oracle 10g, DB2, Netezza, Linux, SQL, Hadoop, AWS, SPSS, Spark, Pig, Hive, Windows 7, MS Excel, IBM Cognos, Excel, VBA, R, Shiny, Python, IBM Jazz, HP Quality Center.

**Client: NBT Bank, Norwich, NY Nov 14 – Oct 15**

**Role: Data Scientist**

**Description:** NBTBank is an independent community bank based in Norwich, NY. The project involved working with end users and analysts to provide analysis of credit card customer base on demographical basis and forecasting risk. The permissions to grant a credit card to an individual were based on the analysis done. The information of the person applying for the credit card is gathered and processed for further approval or rejection.

**Responsibilities:**

* Involved in Analysis & Marketing Team to make business decisions
* Involved with key departments to analyze areas and discuss the primary model requirements for the project
* Documented methodology, data reports and machine learning model results and communicated with the Project Team / Manager to share the knowledge
* Performed competitor and customer analysis, risk and pricing analysis and forecasted results for credit card holders on demographical basis
* Used Machine Learning algorithms and Natural Language Processing(NLP) for response modeling and fraud detection efforts for Credit cards
* Developed needs-based segmentation that aided management in gaining a deeper understanding of consumer behavior. These segments assisted management in development and marketing of credit cards.
* Performed machine learning to estimate the probability of a new customer being classified as a good or bad customer.
* Design, develop and produce reports that connect quantitative data to insights that drive and change business.
* Supported client by developing Machine Learning Algorithms on big data using PySpark to analyze transaction fraud, Cluster Analysis etc.
* Perform ad hoc custom analysis as needed using SQL and R.
* Designed and published visually rich and intuitively interactive Tableau workbooks and dashboards for executive decision making.
* Maintain and enhance data model with changes and furnish with definitions, notes, reference values and check lists.

**Environment:** DB2, Teradata, Hadoop, Spark, SPSS, Pig, Hive, XML, Excel, MS Access, VBA, Tableau, Python, R, Linux, SharePoint, HP Quality Center.

**Client: Michaels Stores, Irving, TX Jan 14 – Oct 14**

**Role: Data Scientist**

**Description:** Michaels Stores, Inc. is an American arts and crafts retail chain. It currently operates more than 1262 stores (consisting of 1,145 Michaels stores in 49 US states and Canada, 118 Aaron Brothers stores and 34 Pat Catan's stores). The project was in the processing of demographic and clickstream data of customers in identifying potential customers based on statistical analysis.

**Responsibilities:**

* Supported the assigned project manager and lead data scientist with creating detailed project plans and assisting in developing, scheduling and tracking project timelines.
* Conducted RFM analysis for a client and split the customer database into different RFM segments, which were then used in Direct Marketing Campaign.
* Developed test plan for conducting Direct Marketing Campaign and analyzed the outcome of it and suggested the desired direction the client should proceed.
* Evaluated overall marketing campaign performance and success in short/long term periods (post-campaign analysis).
* Created campaign report by comparing test and control groups’ Key Performance Indicators (KPIs) such as Average Sales, Click-through Response Rate, Success Rate and ROI.
* Built machine learning regression models to predict sales using python
* Created web application using R shiny for project stakeholders.
* Recommended solutions to improve future campaign performance
* Built machine learning algorithms to empower sales teams in the field by identifying and prioritizing leads for Cross-Selling different products
* Developed customer profiling and analyzed their shopping pattern across different channels.
* Performed different segmentation methodologies and came up with Best Customer Segments and Product affinity groups for the clients’ database.
* **Created Dashboards to present critical** company data in a single report.

**Environment:** DB2, Netezza, SQL Server, Python, R, Shiny, SPSS, Hadoop, Tableau, MS Office Suite (Word, Excel, Access and PowerPoint), VBA, PeopleSoft, HP Quality Center.

**ConAgra Foods, Omaha, NE Jan 13 - Jan 14**

**Role: Data Analyst/ Scientist**

**Description:** ConAgra Brands is an American packaged foods company. ConAgra makes and sells products under various brand names that are available in supermarkets, restaurants, and food service establishments. The work involved various projects such as Sales & Marketing analytics, clustering and Resource optimization etc.

**Responsibilities:**

* Build response modeling for customer base and designed a target marketing plan for ConAgra Foods peanut butter brand
* Build machine learning regression model to identify brands which have cross price elastic effect on competing brands.
* Identified most valuable potential customers for promoted products such as yogurt and organic food, through a series of consumer behavior analyses: RFM analysis, Market Basket and Clustering analysis for increasing response rate, sales and profits.
* Developed machine learning solutions to optimize loyalty/retention campaigns, by applying customer behavior consistency and designing sequential offers to attract customers to higher levels of segments
* Fulfilled A/B tests to further improve marketing campaign response rate and effectiveness
* Performed Competitor analysis and studied the impact of deal and price on brand’s market growth.
* Prepared scripts to ensure proper data access, manipulation and reporting functions with R programming languages.
* Extensively used Tableau for generating compelling visuals, Microsoft Word, Excel and Power point for creating presentations and quality reports

**Environment:** SQL, DB2, Netezza, Linux, Tableau, MS OFFICE, WINDOWS, MS Word, MS Excel, VBA, Python, R, MS Access, MS PowerPoint, MS Project.

**Client: Development Credit Bank Limited, Kolkata, India Jun 09 – Jul 11**

**Role: Quantitative Analyst**

**Description:** Development Credit Bank Limited (DCB) is a national banking which provides a full range of commercial and retail banking services with branches all over India. DCB has consumer lending, commercial lending and investment management as business segments. The project was in the processing of demographic data of customers in identifying potential customers based on statistical analysis.

**Responsibilities:**

* Created models for time-series forecasting, multi-variate analysis using R platform.
* Displayed strong excel skills using VLOOKUP, HLOOKUP, SUMPRODUCT, Index match functions in addition to generating pivot tables and central tendencies such as mean, median etc.
* Combined data from many tables using SQL and transformed it.
* Performed statistical analysis on the data for demographic spread of the customers and identified the possible locations for new branch.
* Developed final reports with graphs and charts to showcase the findings to the management.
* Displayed excellent communication skills in handling ad-hoc tasks though email and telephone
* Completed tasks well ahead of deadline to enable Client ample time for review.
* Worked without supervision after 3 months of initial career and was handling a one-on-one basis with Client
* Generated and updated Monthly and weekly data packages covering metrics for various countries such as GDP, Unemployment Rate, PPP and visualizing the y-o-y and m-o-m growth
* Extracted data from SQL servers in Excel format.

**Environment:** MS SQL Server 2000, T-SQL, R, MS Excel, MS Access

**Education:**

- MS in Information Systems Technologies in 2013 from Wilmington University, DE, USA

- M.Sc. in Computer Science in 2009 from University of Calcutta, India

- B.Sc. in Computer Science in 2007 from University of Calcutta, India