**Maddinapudi Ravi**





**Data Science | Machine Learning | R & Python Programming | Cloud**

**maddinapudih@gmail.com | (973)531-7069**

A career minded professional with 8+ years of IT experience includes in **Data Science** (Machine Learning, Deep Learning, Text Mining), Data/Business Analytics, Data Visualization, Data Warehousing, Data Governance & Operations.

**SUMMARY:**

* Extensive experience in Text Analytics, developing different Statistical **Machine Learning, Data Mining** solutions to various business problems and generating data visualizations using **R, Python** and **Tableau**.
* Expertise in transforming business requirements into **analytical models, designing algorithms, building models, developing data mining** and **reporting solutions** that scales across massive volume of **structured** and **unstructured data**.
* Equipped with experience in utilizing statistical techniques which include Correlation, Hypotheses modelling, Inferential Statistics as well as data mining and modelling techniques using **Linear and Logistic regression, clustering, decision trees, and k-mean clustering**.
* Documenting new data to help source to target mapping. Also updating the documentation for existing **data assisting** with **data profiling** to maintain **data sanitation, validation**.
* Expertise in implementing scalable **Statistical** & **Predictive Decision Science Models** using **Machine Learning** platforms like **R** & **Python Data Science Packages** (Scikit-Learn, Pandas, NumPy, Spark MLib).
* Identifies what data is available and relevant, including internal and external data sources, leveraging new data collection processes such geo-location information.
* Proficient in research of current process and emerging technologies which need analytic models, data inputs and output, **analytic metrics** and **user interface** needs.
* Expertise in building Supervised and Unsupervised **Machine Learning** experiments using **Azure** & **Amazon** utilizing multiple algorithms to perform detailed **predictive analytics** and **building Web Services** models for all types of data: continuous, nominal, and ordinal.
* Understanding on **Hadoop MapReduce** & **Amazon EMR** or any big data frameworks.
* Mitigated risk factors through careful analysis of financial and statistical data. Transformed and processed raw data for further **analysis, visualization, and modelling.**
* Team builder with excellent communications, time & resource management & continuous client relationship development skills.

**Certifications:**

* **Microsoft Certification in Essentials of Data Science**
* **Amazon Web Services Solution Architect Associate Level**

**Programming & Scripting:**

* R Programming & R Studio
* Python
* SQL
* Bash

**Development Tools:**

* Azure & Amazon Machine Learning
* Talend, Tableau, Gretl
* Notepad++
* Jupyter, Spyder
* Google, Elastic Search Analytics
* HDFS, Hive, Spark, Kafka
* SQL Server 2008

**Techniques & Algorithms:**

* Machine Learning
* Regression
* K-Means Clustering
* Data Mining & Cleaning
* Statistics
* Decision Trees
* Random Forest

**Databases:**

* Oracle
* SQL/MySQL
* Cassandra
* MongoDB

**Project Life Cycle:**

* Data Analysis
* Business Analysis & Monitoring
* Project Scheduling
* Deployment
* Operationalization

**PROFESSIONAL EXPERIENCE:**

**Data Scientist**: **June 2016 – Present**

**McAfee, Plano, TX.**

Roles & Responsibilities:

* Designed applications of **Machine learning, Statistical Analysis** and **Data visualizations** with challenging large data processing problems.
* Involved writing the mapping specifications for converting the legacy building and warehouse datasets.
* Worked with various databases like **Oracle, SQL** and performed the computations, log transformations, feature engineering, and Data exploration to identify the insights and conclusions from complex data using **R- programming** in **R-studio**.
* Implemented predictive models using machine learning algorithms **linear regression** and **linear boosting algorithms** and performed in- depth analysis on the structure of models, compared the performance of all the models and found tree boosting is the best for the prediction.
* Applied concepts of **R-squared, R.M.S.E, P-value** in the evaluation stage to extract interesting findings through comparisons.
* Performed in-depth statistical analysis and data mining methods using **R**, including **Cluster analysis, Logistic Regression,** and **boosting models**.
* Proficient in the entire **CRISP-DM** life cycle and actively involved in all the phases of project life cycle including **data acquisition, data cleaning, data engineering.**
* Extensively used **Azure Machine Learning** to set up the experiments and creating Web services for the **predictive analytics.**
* Integrated SAS datasets into Excel using **Dynamic Data Exchange**, using SAS to analyze data, statistical tables, listings and graphs for reports.
* Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of **Spark**.
* Experienced in handling large datasets using Partitions, **Spark** in Memory capabilities, Broadcasts in Spark, Effective & efficient Joins, Transformations and other during ingestion process itself.
* Worked on writing complex **SQL** queries in performing Data analysis using window functions, joins, improving performance by creating partitioned tables.
* Developed SQL procedures to synchronize the dynamic data generated from **GTID systems** with **the Azure SQL Server.**
* Prepared multiple dashboards using **Tableau** to reflect the data behavior over period of time Analyzed and worked with all aspects of **regression models** (OLS etc.)
* Responsible for working with stakeholders to troubleshoot issues, communicate to team members, leadership and stakeholders on findings to ensure models are well understood and optimized.

**Sr Data Analyst: July 2015 – April 2016**

**Bed, Bath & Beyond Inc, Union, NJ.**

Roles & Responsibilities:

* Setup storage and data analysis tools in **Amazon Web Services cloud computing infrastructure**.
* Used **pandas, numpy, Seaborn, scipy, matplotlib, sci-kit-learn** in Python for developing various **machine learning algorithms.**
* Experience with **NoSQL databases**, such as **MongoDB, Cassandra, HBase** and Utilized SQL, NoSQL databases, Python programing and API interaction.
* Experience using **ETL** and **data visualization tools**.
* Implemented Classification using supervised algorithms like **Logistic Regression, Decision trees, Naive Bayes**.
* Updated Python scripts to match training data with our database stored in **AWS Cloud Search**, so that we would be able to assign each document a response label for further classification.
* Data transformation from various resources, data organization, features extraction from raw and stored.
* Handled importing data from various data sources, performed transformations **using Hive, MapReduce**, and loaded data into **HDFS**.
* Deploy configurations of the **Cloudera Distribution** of **Hadoop** from both the command line and Cloudera Manager.
* Apply quality assurance best practices for predictive modeling/analytics services.
* Involved in defining the source to target **data mappings, business rules**, and **data definitions**.
* Performed automation engineer tasks and implemented the **ELK stack** (Elasticsearch, Kibana) for AWS EC2 hosts.
* Defining the list codes and code conversions between the source systems and the data mart using **Reference Data Management (RDM).**
* Designed Cloud Architecture documents and developed cloud infrastructure for a compute environment utilizing **Amazon Web Services** (Hands on experience designing and deploying AWS services: VPCs, EC2s, S3, RedShift, MySQL DBs, Snowball, ELB, Auto-scaling, IAM policies/roles and Security and Networking Services on Linux and Windows OS).
* Performing an end to end **Informatica ETL Testing** for these custom tables by writing complex SQL Queries on the source database and comparing the results against the target database.
* Extracting the source data from **Oracle tables, MS SQL Server, sequential files and excel sheets**.
* Developing and maintaining Data Dictionary to create metadata reports for technical and business purpose.
* Build and maintain dashboard and reporting based on the statistical models to identify and **track key metrics** and **risk indicators**.
* Parse and manipulate raw, complex data streams to prepare for loading into an analytical tool.
* Broad knowledge of programming, and scripting (especially in **R / Python**).
* Involved in developing Patches & Updates Module.

**Data Analyst: July 2012 – December 2014**

**HSBC, Hyderabad, IN**

Roles & Responsibilities:

* Responsible for design and development of **R/Python** programs to prepare transform and harmonize data sets in preparation for modeling.
* Developed large data sets from **structured and unstructured data**. Perform **data mining**.
* Partnered with modelers to develop data frame requirements for projects.
* Performed **Ad-hoc reporting/customer profiling, segmentation** using R/Python.
* Tracked various campaigns, generating customer profiling analysis and data manipulation.
* Provided R/SQL programming, with detailed direction, in the execution of **data analysis** that contributed to the final project deliverables. Responsible for **data mining**.
* Analyzed large datasets to answer business questions by generating reports and outcome.
* Worked in a team of programmers and data analysts to develop insightful deliverables that support data-driven marketing strategies.
* Involved in fixing bugs and minor enhancements for the front-end modules.
* Implemented **Microsoft Visio** and Rational Rose for designing the Use **Case Diagrams, Class model, Sequence diagrams,** and **Activity diagrams** for **SDLC** process of the application.
* Maintenance in the testing team for **System testing/Integration/UAT**.
* Guaranteeing quality in the deliverables.
* Conducted Design reviews and Technical reviews with other project stakeholders.
* Was a part of the complete life cycle of the project from the requirements to the production support.
* Implemented the project in **Linux environment**.
* Involved in loading data from **RDBMS** and **web logs** into **HDFS**.
* Worked on loading the data from **MySQL** to **HBase** where necessary using **Sqoop**.
* Developed **Hive** queries for Analysis across different banners.
* Launching **Amazon EC2 Cloud Instances** using Amazon Images (Linux/ Ubuntu) and Configuring launched instances with respect to specific applications.
* Exported the result set from **Hive** to **MySQL** using **Sqoop** after processing the data.
* Performed performance improvement of the existing **Data warehouse** applications to increase efficiency of the existing system.

**Jr Data Analyst: May 2009 – June 2012**

**Dr Reddy’s Lab, Hyderabad, IN**

Roles & Responsibilities:

* Performed **data analysis** and **data profiling** using complex SQL on various sources systems including **Teradata, SQL Server**.
* Developed the **logical data models** and **physical data models** that confine existing **condition/potential status data** fundamentals and data flows using **ER Studio**.
* Involved in analysis of Business requirement, Design and Development of **High level** and **Low-level designs**, **Unit** and **Integration testing**.
* Implementation of **Metadata Repository**, Maintaining **Data Quality, Data Cleanup procedures, Transformations, Data Standards, Data Governance program, Scripts, Stored Procedures, triggers** and **execution of test plans**.
* Reviewed and implemented the naming standards for the entities, attributes, alternate keys, and primary keys for the **logical model**.
* **Web crawling** and **text mining** techniques to score referral domains, generate keyword taxonomies, and assess commercial value of bid keywords.
* Coordinated meetings with vendors to define requirements and system interaction agreement documentation between **client** **and vendor system**.
* Data analysis may encompass a range of analytical capabilities including **aggregations, statistical analysis, regression, time-series analysis, clustering, network analysis** and other **statistical modeling** techniques.
* Creation of multimillion bid keyword lists using extensive web crawling. Identification of metrics to measure the quality of each list (yield or coverage, volume, and keyword average financial value).
* **Processing, cleansing,** and **verifying** the integrity of data used for analysis.
* Enterprise Metadata Library with any changes or updates.
* Document **data quality** and **traceability** documents for each source interface.
* Worked on DTS Packages, DTS Import/Export for transferring data between SQL Server 2000 to 2005.
* Generate weekly and monthly asset inventory reports.

**EDUCATION:**

* **Master’s** in Computer Engineering.
* **Bachelor’s** in Electronics and Communication Engineering.