Data Scientist Consultant Data Scientist Consultant Data scientist Birmingham, AL   8+ years of IT experience which includes Machine Learning, Data mining with large datasets of Structured and Unstructured data, Data Acquisition, Data Validation, Predictive modeling, Data Visualization.   A deep understanding of Statistical Modeling, Multivariate Analysis, Big data analytics and Standard Procedures Highly efficient in Dimensionality Reduction methods such as PCA (Principal component Analysis), Factor Analysis etc. Implemented bootstrapping methods such as Random Forests (Classification), K-Means Clustering, KNN, NaiveBayes, SVM, Decision Tree, BFS, Linear and Logistic Regression Methods.   The experience of working in text understanding, classification, pattern recognition, recommendation systems, targeting systems and ranking systems using Python.

  Experience with Natural Language Processing (NLP).   Proficiency in application of statistical prediction modeling, machine learning classification techniques and econometric forecasting techniques.   In-depth knowledge of statistical procedures that are applied in Supervised / Unsupervised problems.   Experience in the application of Neural Network, Support Vector Machines (SVM), and Random Forest.   Experienced in working with advanced analytical teams to design, build, validate and refresh data models that enable the next generation of sophisticated solutions for global clients.   Extensively worked on Python 3.5/2.7 (NumPy, Pandas, Matplotlib, NLTK and Sci-kit learn).   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.   Experience in implementing data analysis with various analytic tools, such as Anaconda 4.0 Jupyter Notebook 4.X, R 3.0(ggplot2) and Excel. Experience in designing star schema, Snowflake schema for Data Warehouse, ODS architecture. Hands on experience in business understanding, data understanding, and preparation of large databases.   Experience in working with relational databases (MySQL, Oracle) with advanced SQL programming skills   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.   Experience in using various packages in Rand python-like ggplot2, caret, dplyr, Rweka, gmodels, RCurl, tm, C50, twitter, NLP, Reshape2,

rjson, dplyr, pandas, NumPy, Seaborn, SciPy, Matplotlib, Scikit-learn, Beautiful Soup, Rpy2. Extensive experience in Text Analytics, generating data visualizations using R, Python and creating dashboards using tools like Tableau.   Hands on experience with big data tools like Hadoop, Spark, Hive, Pig, Impala, PySpark, Spark SQL.   Experience in using one or more cloud computing frameworks, such as AWS, Azure, Google Cloud, etc.   Mapping and tracing data from system to system in order to establish data hierarchy and lineage.   Experience with distributed data/computing tools, Map/Reduce, Hadoop, Hive, Spark, MySQL.   Worked on Tableau to create dashboards and visualizations.   Proficiency in various type of optimization, Market Mix modeling, Segmentation, Time Series, Price Promo models etc.   Identifies/creates the appropriate algorithm to discover patterns, validate their findings using an experimental and iterative approach.   Strong skills in statistical methodologies such as A/B test,Experiment design, Hypothesis test, ANOVA, Crosstabs, Ttests and Correlation Techniques.   Applies advanced statistical and predictive modeling techniques to build, maintain, and improve on multiple real-time decision systems. Experience in designing stunning visualizations using Tableau software and publishing and presenting dashboards, Storyline on web and desktop platforms.   Proficiency in SAS (Base SAS, Enterprise Guide, Enterprise Miner)   Excellent communication skills (verbal and written) to communicate with clients and team, prepare deliver effective presentations.   Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design Specification and Testing as per Cycle in both Waterfall and Agile methodologies.   Strong experience in interacting with stakeholders/customers, gathering requirements through interviews, workshops, and existing system documentation or procedures, defining business processes, identifying and analyzing risks using appropriate templates and analysis tools.   Closely works with product managers, Service development managers, and product development team in productizing the algorithms developed Work Experience Data Scientist Consultant BBVA Compass - Birmingham, AL August 2018 to Present Description: BBVA Compass understands that every individual and company has unique dreams and ambitions, needs and wants. We realize that few take the same path in the faster, busier, and more complex world we live in. We get it. Whichever path you choose,

and whenever you need us, we want to create opportunities for your bright future. From the smallest moment, to the largest personal or professional life event, BBVA Compass is there for you. Responsibilities: Performed Data Profiling to learn about behavior with various features such as traffic pattern, location, and time, Date and Time etc. Extensively used Python's multiple data science packages like Pandas, NumPy, Matplotlib, SciPy, Sci-kit learn and NLTK. Worked on data cleaning, data preparation and feature engineering with Python 3.X. In depth understanding of NLP for tokenization, lemmatization and stemming using NLTK, Spacy, Pattern libraries. Experience in evaluating and proposing solutions to NLP problems through various approaches. Deep background in information retrieval (OCR, Speech-to-Text etc.), Natural Language Processing (NLP), knowledge representation or computational linguistics. Understanding of NLP techniques for text representation, semantic extraction techniques, data structures and modeling. Experience in Java, Python and NLP/MLframeworks and libraries. Worked to apply a broad array of capabilities spanning machine learning, statistics, text-mining/NLP, and modeling to extract insights to structured and unstructured healthcare data sources, pre-clinical, clinical trial and complementary real world information streams to apply a broad array of capabilities spanning machine learning, statistics, text-mining/NLP, and modeling to extract insights to structured and unstructured healthcare data sources, pre-clinical, clinical trial and complementary real world information streams. Developed User Defined Functions in Python for rapid analysis and performed data imputation using Sci-kit learnpackage in Python. Developed the applications, models, used appropriate algorithms for arriving at the required insights by analyzing business requirements. Worked with unsupervised (K-means, DBSCAN) and supervised learning techniques (Regression, Classification) for feature engineering and did Principal Component Analysis for dimensionality reduction of features. Used the Classification machine learning algorithms Na ve Bayes, Linear regression, Logistic regression, SVM, Neural Networks and used Clustering Algorithm K Means. Performed text classification task using NLTK package and implemented various natural language processing techniques. Worked collaboratively with senior management to develop strategy and approach to defining business challenges to be answered by data science. Established partnerships

with product and engineering teams and work closely with other teams. Used Spark-StreamingAPIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time. Worked on end to end pipe line in Spark and on Apache spark for analyzing the live streaming data. Worked on Spark Python modules for machine learning and predictive analytics in Spark on AWS. Explored and analyzed the customer specific features by using SparkSQL. Created Hive scripts to create external, internal data tables on Hive. Worked on creating datasets to load data into HIVE. Worked on Tableau for data visualization to create reports, dashboards for insights and business process improvement. Created the dashboards and reports in tableau for visualizing the data in required format. Environment: Spark, Apache Spark, Hive, Machine learning, Python, NumPy, NLTK, Pandas, SciPy, MySQL, Tableau, Sqoop, HBase, HDFS, Tableau, DynamoDB, Mongo DB, SQL Server, and ETL. Data Scientist Visa - Austin, TX May 2017 to July 2018 Description: Visa is a dynamic, global enterprise, and innovation is at the heart of everything we do. We're looking for smart, ambitious people to join our Austin team. We are a global payments technology company working to enable consumers, businesses, banks and governments to use digital currency. Responsibilities: Analyzed Pandas, NumPy, seaborn, SciPy, Matplotlib, Scikit-learn, NLTK in Python for developing various machine learning algorithms and utilized machine learning algorithms such as linear regression, multivariate regression, naive Bayes, Random Forests, K-means, &KNN for data analysis. Demonstrated experience in design and implementation of Statistical models, Predictive models, enterprise data model, metadata solution and data life cycle management in both RDBMS, Big Data environments. Utilized domain knowledge and application portfolio knowledge to play a key role in defining the future state of large, business technology programs. Developed MapReduce/SparkPython modules for machine learning& predictive analytics in Hadoop on AWS. Implemented a Python-based distributed random forest via Python streaming. Performed Source System Analysis, database design, data modeling for the warehouse layer using MLDM concepts and package layer using Dimensional modeling. Created ecosystem models (e.g. conceptual, logical, physical, canonical) that are required for supporting services within the enterprise data

architecture (conceptual data model for defining the major subject areas used, ecosystem logical model for defining standard business meaning for entities and fields, and an ecosystem canonical model for defining the standard messages and formats to be used in data integration services throughout the ecosystem)   Developed LINUXShell scripts by using NZSQL/NZLOAD utilities to load data from flat files to Netezza database.   Designed and implemented system architecture for AmazonEC2 based cloud-hosted solution for client.   Tested Complex ETL Mappings and Sessions based on business user requirements and business rules to load data from source flat files and RDBMS tables to target tables.   Hands on database design, relational integrity constraints, OLAP, OLTP, Cubes and Normalization (3NF) and De-normalization of database.   Developed MapReduce/Spark Python modules for machine learning& predictive analytics in Hadoop on AWS. Worked on customer segmentation using an unsupervised learning technique - clustering.   Worked with various Teradata15 tools and utilities like TeradataViewpoint, Multi Load, ARC, TeradataAdministrator, BTEQ and other Teradata Utilities.   Utilized Spark, Scala, Hadoop, HBase, Kafka, SparkStreaming, MLlib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction.   Analyzed large data sets apply machine learning techniques and develop predictive models, statistical models and developing and enhancing statistical models by leveraging best-in-class modeling techniques.   Environment: Erwin r9.6, Python, SQL, Oracle 12c, Netezza, SQL Server, Informatica, Java, SSRS, PL/SQL, T-SQL, Tableau, MLlib, regression, Cluster analysis, Scala NLP, Cassandra, Map Reduce, Spark, Kafka, Mongo DB, logistic regression, Hadoop, Hive, Teradata0, random forest, OLAP, Azure, Maria DB, SAP CRM, HDFS, ODS, NLTK, SVM, JSON, Tableau, XML, AWS. Data Scientist Black Tree Healthcare Consulting - King of Prussia, PA January 2016 to April 2017 Description: Black Tree Healthcare Consulting provides revenue cycle, outsourcing and clinical consulting services to the home health, hospice and skilled nursing industries. The primary goal of the project was to determine the most efficient input-output ratio through data modeling and analysis, thus providing the best resource combination.   Responsibilities: Incident and problem management, coordinating resolution of data movement disruptions.   Capacity analysis and planning for a sizable

application server environment. Performance metering and tuning for application servers, database, monitors and alerts.   Risk identification and evaluation, managing and improving upon internal controls which mitigate risks   Working with development teams to identify enhancement opportunities and resolve code defects   Code deployment and configuration management, Coordination and execution of changes within a complex testing environment   Data collection procedures enhancement to include information that is relevant for building analytic systems processing, cleansing, and verifying the integrity of data used for analysis.   Building analytic models using a variety of techniques such as logistic regression, risk scorecards and pattern recognition technologies.   Analyze and understand large amounts of data to determine suitability for use in models and then work to segment the data, create variables, build models and test those models.   Excellent understanding of machine learning techniques and algorithms, such as Logistic Regression, SVM, Random Forests, Deep Learning etc.   In depth understanding and experience in NLP and Deep learning.   Expertise in Sentiment Analysis, Entity Extraction, Document Classification, Topic Modeling, Natural Language Understanding (NLU) and Natural Language Generation (NLG).   Leads all NLP/NLG driven solutions for the project   Design NLP models for searching structured/unstructured data in real/near-real time   Interacts regularly with product team members ensuring successful integration of NLP solutions in the product architecture   Keeps the solutions updated with the recent developments in NLP/NLG.   Perform a proper EDA, Univariateand bi-variate analysis to understand the intrinsic effect/combined.   Performing ad-hoc analysis and presenting results in a clear manner and constant tracking of model performance. Worked with Data governance, Data quality, Data lineage, Data architect to design various models. Designed data models and data flow diagrams using Erwin and MS Visio. Independently coded new programs and designed Tables to load and test the program effectively for the given POC's. Developed Implemented & maintained the Conceptual, Logical & Physical Data Models using Erwin for Forward/Reverse Engineered Databases.   Experience with common data science toolkits, such as R, Python, Spark, etc.   Applied statistics skills, such as statistical sampling, testing, regression, etc. Work with technical and development teams to deploy models. Build Model Performance

Reports and   Designing Technical Documentation to support each of the models for the product line.   Exploratory Data Analysis and Data Visualizations using R, and Tableau. Established Data architecture strategy, best practices, standards, and roadmaps.   Lead the development and presentation of a data analytics data-hub prototype with the help of the other members of the emerging solutions team.   Involved in analysis of Business requirement, Design and Development of High level and Low-level designs, Unit and Integration testing. Interacted with the other departments to understand and identify data needs and requirements.   Worked with several R packages including knitr, dplyr, Spark-R, Causal Infer, space time.   Environment: R 3.x, UNIX, Python 3.5.2, MLLib, SAS, regression, logistic regression, Hadoop, NoSQL, MySQL, OLTP, Random forest, OLAP, HDFS, ODS. Data Analyst Grey Hound - Dallas, TX March 2014 to December 2015 Description: Greyhound is an intercity bus common carrier serving over 3,800 destinations across North America. Behavior of users on our website generate large amount of data, which challenges our ability to gather, parse, analyze and infer from it. So, a major task is to implement new capabilities for reporting from existing data warehouse and empower users with better solutions to increase operational efficiency for data warehousing needs. The aim of the project is to gain more visibility into key metrics of operations in the company and to gain insights into how to drive improvements and take the business to a whole new level with advanced analysis through the application of data science.   Responsibilities: Involved in requirement gathering, data analysis and Interacted with Business users to understand the reporting requirements, analyzing BI needs for the user community.   Created Entity/Relationship Diagrams, grouped and created the tables, validated the data, identified PKs for lookup tables.   Involved in modeling (Star Schema methodologies) in building and designing the logical data model into Dimensional Models.   Created and maintained logical, dimensional data models for different Claim types and HIPAA Standards. Implemented one-many, many-many Entity relationships in the data modeling of Data warehouse. Experience working with MDM team with various business operations involved within the organization.   Identify the Primary Key, Foreign Key relationships across the entities and across subject areas.   Developed ETL routines using SSIS packages, to plan an effective package

development process and design the control flow within the packages.   Worked with Big Data Architects for setting up Big Data Platform in the organization and on Hive platform to create Hive Data Models   Developed customized training documentation based on each client's technical needs and built a curriculum to help each client learn both basic and advanced techniques for using Postgre SQL.   Took an active role in the design, architecture, and development of user interface objects in Qlik View applications. Connected to various data sources like SQL Server, Oracle, and flat files.   Presented the Dashboard to Business users and cross-functional teams, define KPIs (Key Performance Indicators), and identify data sources.   Designed data flows that (ETL) extract, transform, and load data by optimizing SSIS performance.   Deliver end to end mapping from source (Guide wire application) to target (CDW) and legacy systems coverages to Landing Zone and to Guide wire Reporting Pack.   Involved in loading the data from Source Tables to Operational Data Source tables using Transformation and Cleansing Logic.   Performed the Data Accuracy, Data Analysis, Data Quality checks before and after loading the data.   Resolved the data type inconsistencies between the source systems and the target system using the Mapping Documents.

  Generated tableau dashboards for Claims with forecast and reference lines.   Designed, developed, implemented and maintained InformaticaPower center and Informatica Data Quality (IDQ) application for matching and merging process.   Created ad-hoc reports to users in Tableau by connecting various data sources.   Worked on the reporting requirements for the data warehouse.   Created support documentation and worked closely with production support and testing team.   Environment: Erwin8.2, Oracle 11g, OBIEE, Crystal Reports, Toad, Sybase Power Designer, Datahub, MS Visio, DB2, QlikView 11.6, Informatica. Data Analyst Unisys Global Services - Bengaluru, Karnataka December 2012 to February 2014 Description: Unisys is a global information technology company that builds high-performance, security-centric solutions for the most demanding businesses and governments on Earth. Unisys offerings include security software and services; digital transformation and workplace services; industry applications and services; and innovative software operating environments for high-intensity enterprise computing.   Responsibilities: Implemented and updated analytical methods such as regression modelling, classification tree,

statistical tests and data visualization techniques with Python. Performed exploratory Data Analysis, Data Wrangling and development of algorithms in R and Python for data mining and analysis. Analysis of customer data and other operational data in MySQL and MS Access to provide insights that enable improvements to customer experience. Utilized pandas and NumPy packages in Python to improve data collection and distribution processes as well as to enhance reporting capabilities to provide clear line of sight into key performance trends and metrics. Performed data analysis, data manipulation, data transformation and data mapping of source data from the MySQLserver. Understanding and adherence to the principles of data quality management including metadata, lineage, and business definitions. Analyzed historical demand, filter out outliers/exceptions, identify the most appropriate statistical forecasting algorithm, develop base plan, understand variance, propose improvement opportunities, and incorporate demand signal into forecast and executed data visualization by using plotly package in Python. Participated in all phases of research including data collection, data cleaning, data mining, development models and visualizations. Examine customer feedback and activity for use in detecting or confirming fraud, using a combination of text analytics, statistical modeling, and classification. Deep understanding of Software Development Life Cycle (SDLC) as well as Agile/Scrum methodology to accelerate Software Development iteration. Improvisations and maintenance to existing automated solutions. Used MS Visio, MS Project to assist the team in project planning, quality plan, risk management, requirements management, change management, defect management, change management and release management. Environment: MySQL, Statistical modelling, Python libraries, Pandas, NumPy packages, R, MS Visio, MS Project, MS Access. Data Analyst/ Python Developer Hidden Brains - Hyderabad, Telangana January 2011 to November 2012 Description: Hidden Brains InfoTech Pvt. Ltd is an Enterprise Web & Mobile Apps Development Company. With an industry experience of over a decade, we offer a plethora of client-centric services by enabling customers to achieve competitive advantage through flexible and next generation global delivery models. Responsibilities: Designed and developed transformation logic for BI tools (Informatica) for data transformation into various layers of Data warehouse. Performed code development with

the help of internal Python library that speeds up database querying and allow users to write more resilient ETL jobs. Implemented PostgreSQL, SQL servers to develop stored procedures, views to create result sets to meet varying reporting requirements. Ensured data integrity using advanced excel formulas (lookup functions, pivot table, If Statements etc.) for analyzing data. Hands on experience in writing queries in SQL and R to extract, transform and load (ETL) data from large datasets using Data Staging. Participated in the A/B testing conducted by BIAnalytics team for data extraction and exploratory analysis. Generated dashboards and presented the analysis to researchers explaining insights on the data. Provided analytical insights and decision support tools for executives for accurate decision making by identifying, measuring and recommending improvement strategies for KPIs across all business areas. Submitted summaries, charts, and graphs to team and stakeholders that help to interpret findings based on complex excel reports. Responsible for Credit data related warehouse creation that could help with Risk Assessment for Commercial loans. Performed competitor and customer analysis, risk and pricing analysis and forecasted results for credit card holders on demographical basis. Worked in manipulating various management reports in MS Excel for sales metrics using VLOOKUP and Pivot tables. Involved in estimating, defining, implementing, and utilizing business metrics calculations and methodologies.

Environment: Excel 2010, R, Informatica Power Center 9.0, Python, PostgreSQL, MS SQL Server 200. Education Bachelor's Skills CASSANDRA, HDFS, IMPALA, MAPREDUCE, SQOOP, HBASE, KAFKA, ELASTICSEARCH, ETL, FLUME, HADOOP, INFORMATICA, MONGODB, NLP, REDIS, TABLEAU SERVER, TERADATA, DATA MODELING, DATABASE, DATABASE DESIGN Additional Information Technical Skills Programming & Scripting Languages R, C, C++, JAVA, JCL, COBOL, HTML, CSS, JSP, Java Script Databases SQL, Hive, Impala, Pig, Spark SQL, Databases SQL-Server, My SQL, MS Access, HDFS, HBase, Teradata, Netezza, MongoDB, Cassandra. Statistical Software SPSS, R, SAS. Web Packages ggplot2, caret, dplyr, Rweka, gmodels, RCurl, tm, C50, twitteR, NLP, Reshape2, rjson, plyr, pandas, numPy, seaborn, sciPy, matplot lib, scikit-learn, Beautiful Soup, Rpy2, sqlalchemy. Big data Ecosystem HDFS, PIG, MapReduce, HIVE, SQOOP, FLUME, HBase, Storm, Kafka, Elastic Search, Redis, Flume, Storm, Kafka, Elastic

Search, Redis, Flume, Scoop.   Statistical Methods Time Series, regression models, splines, confidence intervals, principal component analysis and Dimensionality Reduction, bootstrapping   BI Tools Tableau, Tableau server, Tableau Reader, SAP Business Objects, OBIEE, QlikView, SAP Business Intelligence, Amazon Redshift, or Azure Data Warehouse   Database Design Tools and Data Modeling Erwin r 9.6, 9.5, 9.1, 8.x, Rational Rose, ER/Studio, MS Visio, SAP Power designer. Cloud/ ETL Tools AWS, S3, EC2, Informatica Power Centre, SSIS. Operating System Windows, Linux, Unix, Macintosh HD, Red Hat. Big Data / Grid Technologies Cassandra, Coherence, Mongo DB, Zookeeper, Titan, Elasticsearch, Storm, Kafka, Hadoop Tools and Utilities SQL Server Management Studio, SQL Server Enterprise Manager, SQL Server Profiler, Import & Export Wizard, Visual Studio.Net, Microsoft Management Console, Visual Source Safe 6.0, DTS, Crystal Reports, Power Pivot, ProClarity, Microsoft Office, Excel Power Pivot, Excel Data Explorer, Tableau, JIRA,SparkMLlib.   Project:1 Role: Data Scientist

Name: Lee Robbins

Email: hannah01@example.net

Phone: 8162049858