Data Scientist Data Scientist Data Scientist - Nielsen New York, NY ? 8+ years of experience in Machine Learning, Data-mining with large datasets of Structured and Unstructured data, Data Acquisition, Data Validation, Predictive modelling, Data Visualization. ? Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python. ? Data Driven and highly analytical with working knowledge and statistical model approaches and methodologies (Clustering, Regression analysis, Hypothesis testing, Decision trees, Machine learning), rules and ever-evolving regulatory environment. ? Professional working experience in Machine Learning algorithms such as Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, K-Means Clustering and Association Rules. ? Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing data mining and reporting solutions that scale across a massive volume of structured and unstructured data. ? Experience with data visualization using tools like Ggplot, Matplotlib, Seaborn, Tableau and using Tableau software to publish and presenting dashboards, storyline on web and desktop platforms. ? Experienced in python data manipulation for loading and extraction as well as with python libraries such as NumPy, SciPy and Pandas for data analysis and numerical computations. ? Well experienced in Normalization, De-Normalization and Standardization techniques for optimal performance in relational and dimensional database environments. ? Experience in multiple software tools and languages to provide data-driven analytical solutions to decision makers or research teams. ? Familiar with predictive models using numeric and classification prediction algorithms like support vector machines and neural networks, and ensemble methods like bagging, boosting and random forest to improve the efficiency of the predictive model. ? Worked on Text Mining and Sentimental analysis for extracting the unstructured data from various social Media platforms like Facebook, Twitter, and Reddit. ? Develop, maintain and teach new tools and methodologies related to data science and high-performance computing. ? Extensive hands-on experience and high proficiency with structures, semi-structured and unstructured data, using a broad range of data science programming languages and big data tools including R, Python, Spark, SQL, Scikit Learn, Hadoop

MapReduce ? Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design Specification and Testing as per Cycle in both Waterfall and Agile methodologies. ? Adept in statistical programming languages like R and Python including Big Data technologies like Hadoop, Hive. ? Hands on experience with RStudio for doing data pre-processing and building machine learning algorithms on different datasets. ? Collaborated with the lead Data Architect to model the Data warehouse in accordance with FSLDM subject areas, 3NF format, Snowflake schema. ? Worked and extracted data from various database sources like Oracle, SQL Server, and DB2. ? Implemented machine learning algorithms on large datasets to understand hidden patterns and capture insights. ? Predictive Modelling Algorithms: Logistic Regression, Linear Regression, Decision Trees, K-Nearest Neighbours, Bootstrap Aggregation (Bagging), Naive Bayes Classifier, Random Forests, Boosting, Support Vector Machines. ? Flexible with Unix/Linux and Windows Environments, working with Operating Systems like Centos5/6, Ubuntu13/14, Cosmos. Authorized to work in the US for any employer Work Experience Data Scientist Nielsen - New York, NY August 2018 to Present Nielsen Company is the leading provider of entertainment metadata and media recognition technology that powers discovery features and discover the music, TV shows, movies and sports they love across the world's most popular entertainment platforms and devices, from Amazon, Apple, Facebook, Google, Time Warner Cable, Tesla and others. Our project deliver mission is using critical data to help our clients (CNN,) grow their business with our extensive and quality data verified by real consumers   Responsibilities: ? Plays a key role in Data Adoption projects from beginning to end; including developing plan, running analyses, summarizing results, and communicating with client ? Consulting with clients to explain methodology, data impacts, and insights along with proper use guidelines and limitations related to new and enhanced services ? Used Pyspark, sparkSQL in making pipelines to extract data coming from MDL in spark environment. ? Used POSTGRE SQL and also worked with Hadoop. ? Develop NLP(Natural Learning Program) and other data mining algorithms to extract useful information from large data sets ? Build data pipelines and Machine Learning(ML) models that run in production in collaboration with software engineers using the tools like Numpy, SciKit ? Used Machine learning concepts-

common families of models, feature engineering & selection, cross-validation and parameter tuning ? Worked with DevOps tools and CI/CD workflows including github, Jenkins, and Docker Swarm ? Used statistical techniques, market research methodologies, research processes, operations to analyze the complexity of consumer businesses, complex analytical challenges and client needs to enable better decisions using the data ? Worked on AWS platform for implementing cloud based solutions. ? Helping to solve client challenges such as performance management, product or methodology evolution POC using the advanced techniques and tools common to the data science world like MLlib and Scikit-learn ? Categorized the clients data and their needs using the Linear regression, Time series Regression, K mean, Neural Networks, Decision trees, Classification.,. ? Uses statistical methodologies to analyze the data using Python ,R, SAS, SPSS, Matlab and to improve the survey quality in production environment ? Executed the company's data mining and modeling activities in support of our clients' online targeting and digital media marketing goals ? Utilizes tools such as Python, Tableau, R etc. to perform complex data analysis and visualizations ? Works closely with internal customers and IT personnel to improve current processes and engineer new methods. This includes support with writing new software, testing and end-user requirements Environment: NLP, Machine learning, Deep Learning, Python, R, Tableau, SAS, MATLAB, AWS, Neural Network Data Scientist Progressive Insurance, OH March 2017 to August 2018 The Progressive Corporation is one of the largest providers of car insurance in the United States. The company also insures motorcycles, boats, RVs and commercial vehicles, and provides home insurance through select companies. Progressive has expanded internationally as well, offering car insurance in Australia. Responsibilities: ? Setup storage and data analysis tools in Amazon Web Services cloud computing infrastructure. ? Used pandas, NumPy, Seaborn, SciPy, matplotlib, sci-kit-learn, NLTK in Python for developing various machine learning algorithms. ? Installed and used CaffeDeep Learning Framework ? Worked on different data formats such as JSON, XML and performed machine learning algorithms in Python. ? Worked as Data Architects and IT Architects to understand the movement of data and its storage and ER Studio 9.7 ? Participated in all phases of datamining; data collection, data cleaning, developing models, validation, visualization and

performed Gap analysis. ? Data Manipulation and Aggregation from a different source using Nexus, Toad, Business Objects, Powerball, and Smart View. ? Implemented Agile Methodology for building an internal application. ? Good knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Secondary Name Node, and Map Reduce concepts. ? As Architect delivered various complex OLAP databases/cubes, scorecards, dashboards and reports. ? Programmed by a utility in Python that used multiple packages (SciPy, NumPy, pandas) ? Implemented Classification using supervised algorithms like Logistic Regression, Decision trees, KNN, Naive Bayes. ? Responsible for design and development of advanced R/ Python programs to prepare to transform and harmonize data sets in preparation for modelling. ? Worked on different data formats such as JSON, XML and performed machine learning algorithms in Python. ? Worked as Data Architects and IT Architects to understand the movement of data and its storage and ER Studio 9.7 ? Participated in all phases of datamining; data collection, data cleaning, developing models, validation, visualization and performed Gap analysis. ? Data Manipulation and Aggregation from a different source using Nexus, Toad, Business Objects, PowerBL,and Smart View. ? 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, Map Reduce, and loaded data into HDFS. ? Interaction with Business Analyst, SMEs, and other Data Architects to understand Business needs and functionality for various project solutions ? Researched, evaluated, architected, and deployed new tools, frameworks, and patterns to build sustainable Big Data platforms for the clients. ? 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. ? Identifying and executing process improvements,

hands-on in various technologies such as Oracle, Informatica, Business Objects. ? Designed both 3NF data models for ODS, OLTP systems and dimensional data models using Star and Snowflake Schemas.   Environment: R 9.0, ODS, OLTP, Bigdata, Oracle 10g, Hive, OLAP, DB2, Metadata, Python, MS Excel, Mainframes MS Vision, Rational Rose. Data Scientist Essendant - Deerfield, IL October 2015 to February 2017 Essendant, formerly known as United Stationers, is a national wholesale distributor of workplace essentials, with consolidated net sales of $5.3 billion. In 2013, it ranked 484 (478 in 2012; 467 in 2011) out of the Fortune 500 companies.   Responsibilities: ? Performed Data Profiling to learn about behavior with various features such as traffic pattern, location, and time, Date and Time etc. ? Application of various machine learning algorithms and statistical modeling like decision trees, regression models, neural networks, SVM, clustering to identify Volume using Scikit-learn package in python, MATLAB. ? Utilized Spark, Scala, Hadoop, HBase, Cassandra, MongoDB, Kafka, Spark Streaming, MLlib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc. and Utilized the engine to increase user lifetime by 45% and triple user conversations for target categories. ? Developed Spark/Scala, Python for regular expression (regex) project in the Hadoop/Hive environment with Linux/Windows for big data resources. Used clustering technique K-Means to identify outliers and to classify unlabeled data. ? Evaluated models using Cross Validation, Log loss function, ROC curves and used AUC for feature selection. ? Developed entire frontend and backend modules using Python on Django Web Framework. ? Addressed overfitting by implementing the algorithm regularization methods like L2 and L1. ? Used Principal Component Analysis in feature engineering to analyze high dimensional data. ? Identified and targeted welfare high-risk groups with Machine learning algorithms. ? Conducted campaigns and run real-time trials to determine what works fast and track the impact of different initiatives. ? Analyzed and calculated the lifetime cost of everyone in the welfare system using 20 years of historical data. ? Performed Multinomial Logistic Regression, Random forest, Decision Tree, SVM to classify package is going to deliver on time for the new route. ? Used MLlib, Spark's Machine learning library to build and evaluate different models. ? Implemented rule-based expertise system from the results of

exploratory analysis and information gathered from the people from different departments. ? Performed Data Cleaning, features scaling, features engineering using pandas and NumPy packages in python. ? Developed Map Reduce pipeline for feature extraction using Hive. ? Created Data Quality Scripts using SQL and Hive to validate successful data load and quality of the data. Created various types of data visualizations using Python and Tableau. Environment: Python 2.x, HDFS, Hadoop 2.3, Hive, Linux, Spark, Tableau Desktop, SQL Server 2012, Microsoft Excel, MATLAB, Spark SQL, PySpark. Data Scientist Coventry Health Care - Downers Grove, IL December 2014 to September 2015 Coventry offers workers' compensation, auto, and disability care- and cost-management solutions for employers, insurance carriers, and third-party administrators. With roots in both clinical and network services, Coventry leverages 30+ years of industry experience, knowledge, and data analytics Responsibilities: ? Responsible for data exploration, cleaning for modeling, participate in model development. ? Used Principal Component Analysis and factor Analysis in feature engineering to analyze high dimensional data in python. ? Performed data cleaning, factorization, feature engineering and feature scaling. ? Visualize, interpret, report findings, and develop strategic uses of data by python libraries like NumPy, Pandas, SciPy, Scikit-Learn. ? Missing value treatment, Outlier capping and anomalies treatment using statistical methods. ? Evaluated models using Cross Validation, Log loss function, ROC Curves and AUC for feature selection. ? Worked with several R packages including GGPLOT, DPLYR, and KNITR. ? Strong skills in data visualization like Matplotlib and Seaborn. ? Created different charts such as Heatmaps, Bar charts, Line charts etc. ? Data mining using the state-of-the-art methods and dimensionality reduction using Principal Component Analysis, t-SNE for visualizing high dimensional data. ? Involved in various pre-processing phases of text data like Tokenizing, Stemming, Lemmatization and converting the raw text data to structured data. ? Participated in all phases of data mining, data collection, data cleaning, developing models, validation, and visualization and performed Gap analysis. ? Constructing the new vocabulary to convert the data into numbers to be processed by the machine by using the approaches like Bag of Words, TF-IDF, Word2vec, and Average Word2Vec. ? Implemented Bi-Directional Recurrent Neural Networks acts

as encoder to process the input and as decoder to generate the output. ? Used Recurrent Neural Networks with LSTM cells to protect the sequence information. ? LSTM cells are implemented in the Recurrent Neural Network to get the longer-term dependencies. ? Used testing methods like A/B Testing, Multi-Variate to measure impact on new initiatives. ? Applied binary classification and parse trees to identify key features of radiology related sentences using Natural language processing (NLP). ? Using NLP developed deep learning algorithms for analyzing text, over their existing dictionary-based approaches. ? Worked on customer segmentation using unsupervised clustering techniques. ? Implemented LSTM layer network of moderate depth to gain the information in the sequence with help of Tensor Flow. ? Created distributed environment of Tensor Flow across multiple devices (CPUs and GPUs) and run them in parallel. ? Implemented machine learning algorithms like Logistic Regression, SoftMax Classifier, Random Forest, Decision Trees. Environment: Cluster Analysis, Regression, Natural Language Processing, Spark ML lib, Logistic regression, SoftMax classifier, Random Forest, Python, SQL, Oracle 12c, NLTK, Recurrent Neural Networks, LSTM cells, Natural Language Toolkit, NumPy, SciPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tensor Flow, Keras. Python Developer Valence Health - Chicago, IL November 2013 to November 2014 Valence Health works with clients to design build and manage value-based care models customized for each client including clinically integrated networks, bundled payments, risk-based contracts, accountable care organizations and provider-sponsored health plans. The project is to create an ETL process and collect data to do analytics and generate reports. Responsibilities: ? Taken part in software development life cycle (SDLC) of the tracking systems Requirements, gathering, Analysis, Detail Design, Development, System Testing and User Acceptance Testing. ? Created UI using HTML, CSS, JavaScript, AJAX, JSON, and JQuery. ? Implemented business logic using PythonWeb frame work Django. ? Designed applications implementing MVC architecture in Pyramid, Zopeframeworks. ? Actively involved in developing the methods for Create, Read, Update and Delete (CRUD) in Active Record. ? Designing mobile search application system requirements and coded back-end and front-end in Python. ? Analysis and Design of application. ? Implemented ModelViewControl architecture in developing web

applications using Django frame work. ? Created backend database T-SQL stored procedures and Jasper Reports. ? Worked with millions of database records on a daily basis, finding common errors and bad data patterns and fixing them. ? Exported/Imported data between different data sources using SQL Server Management Studio. ? Maintained program libraries, users' manuals and technical documentation. ? Managed large datasets using Panda data frames and MySQL. ? Wrote and executed various MYSQL database queries from python using Python-MySQL connector and MySQL dB package. ? Carried out various mathematical operations for calculation purpose using python libraries. ? Built various graphs for business decision making using Python matplotlib library. ? Fetched twitter feeds for certain important keyword using python-twitter library. ? Used Python library BeautifulSoup for web Scrapping. ? Developed applications especially in UNIX environment and familiar with all of its commands. ? Deployed the project into Heroku using GIT version control system. ? Performed troubleshooting, fixed and deployed many Python bug fixes of the two main applications that were a main source of data for both customers and internal customer service team. ? Implement code in Python to retrieve and manipulate data.   Environment: Python 2.7, Django, HTML5/CSS, Pyramid, Zope, MySQL, MS SQL, T-SQL, Jasper Reports, JavaScript, Eclipse, Git, Linux, Shell Scripting. Data Analyst Ediko Systems Inc February 2011 to October 2013 Ediko Systems Integrators, an IBM Premier Business Partner, is a specialist company delivering world-class business solutions leveraging IBM Technologies. EDIKO ensures the delivery of high-quality business integration solutions through the application of sound software architecture principles and using the latest IBM technologies together with agile project management techniques.

  Responsibilities: ? Created new reports based on requirements. ? Responsible for Generating Weekly ad-hoc Reports ? Planned, coordinated, and monitored project levels of performance and activities to ensure project completion in time. ? Automated and scheduled recurring reporting processes using UNIX shell scripting and Teradata utilities such as MLOAD, BTEQ, and Fast Load ? Experience with Perl. ? Worked in a Scrum Agile process & Writing Stories with two-week iterations delivering a product for each iteration ? Worked on transferring the data files to the vendor through sftp&Ftp process ? Involved in defining and Constructing the customer to customer

relationships based on Association with an account & customer ? Created action filters, parameters and calculated sets for preparing dashboards and worksheets in Tableau. ? Experience in performing Tableau administering by using tableau admin commands. ? Worked with architects and, assisting in the development of current and target state enterprise-level data architectures ? Worked with project team representatives to ensure that logical and physical data models were developed in line with corporate standards and guidelines. ? Involved in defining the source to target data mappings, business rules, and data definitions. ? Responsible for defining the key identifiers for each mapping/interface. ? Performed data analysis and data profiling using complex SQL on various sources systems including Oracle and Teradata. ? Migrated three critical reporting systems to Business Objects and Web Intelligence on a Teradata platform ? Created Excel charts and pivot tables for the Adhoc data pull.   Environment: MS Office Suite, MS Visio, MS SharePoint, Test Management Tool, MS Project, Crystal report, HTML. Data Analyst Hidden Brains July 2010 to January 2011 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: ? Processed data received from vendors and loading them into the database. The process was carried out on weekly basis and reports were delivered on a bi-weekly basis. The extracted data had to be checked for integrity. ? Documented requirements and obtained signoffs. ? Coordinated between the Business users and development team in resolving issues. ? Documented data cleansing and data profiling. ? Wrote SQL scripts to meet the business requirement. ? Analyzed views and produced reports. ? Tested cleansed data for integrity and uniqueness. ? Automated the existing system to achieve faster and accurate data loading. ? Generated weekly, bi-weekly reports to be sent to client business team using business objects and documented them too. ? Learned to create Business Process Models. ? Ability to manage multiple projects simultaneously tracking them towards varying timelines effectively through a combination of business and technical skills.   ? Good Understanding of clinical practice management, medical and laboratory billing and insurance claim with processing with process flow

diagrams. ? Assisted QA team in creating test scenarios that cover a day in a life of the patient for Inpatient and Ambulatory workflows.   Environment: SQL, data profiling, data loading, QA team. Education Bachelor's Skills APACHE CASSANDRA (1 year), APACHE HBASE (1 year), ASTERADATA (1 year), Cassandra (1 year), database (5 years), databases (1 year), Excel (5 years), HBase (1 year), Linux (2 years), Matlab (2 years), MongoDB (1 year), MS Office (2 years), MS SQL SERVER (2 years), Python (5 years), Scala (1 year), SQL (7 years), SQL Server (2 years), UNIX (3 years), Visio (2 years), XML (1 year)

Name: Paul Barnes

Email: lbennett@example.org

Phone: (674)872-5558x16015