Jack Maney

Email: jackmaney@gmail.com

Website: jackmaney.com

- Able to derive insights and extract business value from data sets with thousands to billions of rows.
- Writing and optimizing complex SQL queries.
- Building prototypes of data products using relational databases, Apache Spark, MPP systems, and Python.
- Researching and implementing algorithms from white papers and academic literature.
- A/B testing recommender systems.
- Experienced problem-solver with very strong quantitative skills.

Experience

Pinsight Media

Data Scientist III - October 2016--Present

- Built recommender engines for our on-device monetization products. Worked with development teams to ensure that the back end would be set up for AB testing.
- Contributed to a project of predicting demographic attributes for users in our real-time bidding (RTB) system. This resulted in collaborations with Marketing that led to two whitepapers:
    - Guess My Age: Age Discrepancies Across Ad Exchanges
    - There's an App for That: But What Does It Say About My Consumer?
- Mentoring members of the Data Science and Business Intelligence teams.

DST's Applied Analytics Group

Senior Data Scientist - December 2013--October 2016

- Prototyped a product for acquisition of financial advisors.
- Contributed towards an Advisor Segmentation product, including a method of streamlining and summarizing the differences between segments.
- Built a prototype of the Mapper Algorithm (as used in Topological Data Analysis) to better understand high-dimensional data sets. The prototype is written in Python and leverages a Greenplum cluster by way of SQL templates.
- Built prototypes for three components of DST's Predictive Wholesaling product, and assisted the AAG Development team in productionizing the prototypes.

- Created and prototyped a Share Retention metric that provides a measurement of "stickiness" of fund holdings that does not directly depend on price.
- Assisted in building models for a proof of concept for a client.
- Mentored and taught Python to a few members of the Networking team, to facilitate the creation of a Flask web app to automate some types of network change requests.
- Mentoring other members of the Data Science team.

BA Services

Data Scientist - May 2013--November 2013

- Created and cross-validated probit regression models to find the most significant attributes upon which to sort call queues in order to increase customer retention.
- Built an ETL pipeline to import data from a new dialer system.
- Delivered a proposal outlining options for a Data Warehouse solution, including pros and cons of each option.
- Built, Validated, and Deployed business intelligence reports using QlikView.

C2FO

Implementation Specialist (Contract) - January 2013--May 2013

- Optimized the C2FO algorithm for Market Clearing Events, making it run an average of two orders of magnitude faster.
- Organized the restructuring of several KPI business intelligence reports.
- Built, tested, and deployed user management tools for account managers.

Adknowledge

July 2010--November 2012
Titles Held:

- Sr Data Analyst and Mathematician - January 2012--November 2012
- Data Analyst and Mathematician - February 2011--January 2012
- Data Analyst - July 2010-February 2011

- Performed data mining and summarized results that contributed to the winning of a $50,000 advertiser contract
- Presented technical and mathematical concepts to non-technical audiences, including several layers of management and a venture capital investor.

- Developed an application in Perl using DBI for k-means++ clustering. This application is able to handle data sets of millions of rows with 1--100 variables.
- Found a way to implement a regression algorithm--on a dataset with 30 million rows and 250 variables--that was previously thought impossible to implement due to scale
- Built, implemented, deployed, and maintained an ad category recommendation system for advertisers, including developing and measuring performance metrics
- Implemented a genetic algorithm framework to use for behavioral targeting algorithms
- Maintained and documented the ETL pipelines to the Data Analytics team, consisting of over 200 scripts in Perl and Python interfacing with Greenplum, PostgreSQL, Oracle, MySQL, MS SQL, and ActiveMQ
- Refactored and maintained critical business intelligence reports used by machine learning scientists
- Prototyped a flexible, extensible ETL system to reduce a lot of boilerplate code in existing ETL scripts
- Created a web-based data dictionary to store metadata about tables in our warehouse. The front end was written in PHP with SQLite on the back-end to store the metadata.
- Considered a resident expert of our data warehouse
- Contributed to the on-boarding of two interns and two full-time employees

University of South Dakota

Assistant Professor - August 2004--May 2008

- Six peer-reviewed mathematical publications
- Directed two undergraduate Honors Theses and a Master's Thesis in mathematics
- Sole organizer and director of a regional undergraduate mathematics conference
- Taught several courses, including College Algebra, Trigonometry, Calculus (I--III), Foundations of Mathematics, Matrix Theory, and Abstract Algebra
- Served and chaired several committees, including the Curriculum & Instruction committee

Open-source Software

Universal Correlation Coefficient

At the 2011 Joint Statistical Meetings, a paper was presented that introduced the idea of a Universal Correlation Coefficient. This coefficient measures the

degree of dependency (but not the form of dependency) for two discrete random variables.

I have written an R library that implements this Universal Correlation Coefficient. This coefficient can be used to automate the discovery of (potentially non-linear) relationships among pairs of discrete random variables.

pg-utils: Utilities for working with PostgreSQL

Some handy utilities that I've written for processing data in either PostgreSQL or Greenplum.

Python Standard Library List

Lists of names of packages in the Python standard library (for versions 2.6, 2.7, and 3.2-5), along with the code used to grab the list of libraries from the official Python docs. Surprisingly, this is my most popular repository on GitHub.

Diophantus

Diophantus, a pet project created to teach myself Java, originated as Mathematica code that I wrote as a graduate student. The original code generated examples and helped form conjectures for what became a series of two peer-reviewed mathematical publications.

Skills

Languages and Technologies

- Apache Spark, Apache Hive, Apache Hadoop, HDFS
- Python, PySpark, Pandas, NumPy, SciPy, scikit-learn, matplotlib, seaborn, PyCharm
- Relational Databases, Greenplum (Massively Parallel Processing Distributed System), PostgreSQL, Oracle, MySQL, Microsoft SQL Server, MS SQL, OLAP, SQL
- Perl, Moose (OO Perl), DBI, threads, threads::shared, Thread::Queue, Template::Toolkit
- R
- QlikView
- Java, JUnit, Eclipse
- JSON, Parquet
- Git, SVN, GitHub, GitLab, Assembla, Source Control, Version Control
- JIRA, Pivotal Tracker, Confluence

- Linux, openSUSE, Ubuntu, CentOS, RHEL, bash
- The 42matters API for app metadata
- Social Radar API, Facebook Ads API

Other Skills

- Mathematics
- Topological Data Analysis
- Data mining
- Data visualization
- Implementing algorithms and ideas gleaned from academic publications

Education

North Dakota State University

Ph.D. Mathematics, May 2004

B.S Mathematics, Dec 1999

Training Courses and Professional Development

- PostgreSQL Training from Webucator, 2010
- Java Training from Webucator, 2012
- Noble Dialer Operations Training from Noble Systems, 2013
- QlikView Developer Training from Qlik, 2013
- Attended KDD 2014
- Greenplum User Training from Pivotal, 2014
- Data Anonymization Training from Privacy Analytics, 2015
- Hadoop and MapReduce Training from Hortonworks, 2015
- Apache Spark training from Databricks, 2017
- Apache Cassandra training from Learning Tree International, 2017