Steve Petrie

Data Scientist - Bardavon Health Innovations

Overland Park, KS - Email me on Indeed: indeed.com/r/Steve-Petrie/065d2143fde76ee7

A passionate data scientist, whom started his career as a nuclear engineer, the transitioned into the training field, teaching nuclear physics and mathematics to reactor operators. The next 20 years were spent as a database software engineer. The last 4 years, had the opportunity to combine the math and programming skills to focus on predictive analytics, machine learning and artificial intelligence.
Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist

Bardavon Health Innovations - Overland Park, KS - November 2016 to Present

Project: Predict the outcome of patients based upon initial conditions

• Used 7 models: Linear Regression, Non-Linear Regression, Decision tree, Random Forest, Linear SVM, Non-Linear SVM and Neural Net and compared the accuracy of the predictions. The most accurate models were the Decision Tree and Random
Forest at 84% and the least accurate was the Neural Net at 72%. It was decided implement the decision tree since it was the easiest to install in production.
• The implementation of the model was to develop a stored procedure mimicking the if/else statements from the decision tree model. This procedure is then run against all new patients to predict the most probable outcome. The patients with the worst predicted outcome are given the most attention early in the process in order to improve the patient recovery.
• Currently the model is refreshed manually once a month. The plan is to have the model refreshed automatically via python. A tracking system will also be installed to link model versions to predicted patient outcome in order to measure prediction performance.

Technology: Linear Regression, Non-Linear Regression, Decision tree, Random Forest, Linear SVM, Non-Linear SVM and Neural Networks, PCA (Principal Component Analysis), EFA (Exploratory Factor Analysis), Cross-validation, ROC, R, Python Programming (Numpy, Scikit-learn), SQL-Server, SQL

Project: Exploratory Factor Analysis of patient outcome

• The analysis indicated that out of 45 dimensions used only 6 were significant. The intent is to use this information to improve the data collection process by removing unneeded data and using the key dimensions for prediction purposes.

Technology: EFA, PCA, Gini coefficient, R, SQL

Project: Social Network Analysis (SNA) of Seaboard worker comp claims

• The goal was to improve the claim process by analyzing the entities involved in the process.
• The data was represented as a mathematical graph and then the centrality and density were obtained.
• The SNA indicated that out of 130 cases, 3 physical therapists were key to improving the process.

Technology: SNA, R - igraph, SQL, D3

Project: QA Audit improvement - in progress

• Currently the number of patient cases being audited is 100%. Since the number of cases being processed is doubling every 6 months, this soon will be unstainable.
• The goal of the project is to predict which cases most likely will be in error and focus the auditors on those cases.
• The plan is to obtain publicly available data, such as census demographic data, national and state health provider licensing information, plus internal data such as how the provider completes their forms. This information would be used to generate a profile of error prone providers then use cosine similarity to find other providers with similar profiles. At this point training will be provided to those providers to reduce the possible errors.

Technology:: Linear Regression, Non-Linear Regression, Decision tree, Random Forest, Linear SVM, Non-Linear SVM and Neural Networks, PCA, EFA, Cross-validation, ROC, R, Python Programming (Numpy, Scikit-learn), SQL-Server, SQL, Python calls to census demographic data.

Data Scientist / Senior Software Engineer

Jack Henry & Associates - Shawnee, KS - March 2013 to November 2016

Responsibilities:
• Worked using .Net and SQL server on the jhaKnow Data Warehouse product.
• Worked on adding Predictive Analytics to the jhaKnow product

Project: Bank Customer Churn - Used the standard 7 predictive analytics models to predict which customers were most likely to close their account in the next 90 days.

Technology: Linear Regression, Non-Linear Regression, Decision tree, Random Forest, Linear SVM, Non-Linear SVM and Neural Networks, PCA, EFA, Cross-validation, ROC, R, Python Programming (Numpy, Scikit-learn), SQL-Server, SQL

Project: Bank Product Association
• Used R to create a recommendation engine for products sold by a bank. This is same process used by Amazon and NetFlix for recommending products to its customers.
• For the bank in question, it provided insight into how their products relate to each other, thus while product A may not be profitable, it would lead the customer to product B and C, which were profitable.

Technology: Python (numpy, scikit-learn), APIORI, D3, SQL

Project: Bank Market Analysis
• Used a KNN cluster analysis using python of FDIC data combined with Jack Henry product sales data to group banks by how likely they were to purchase a jhaKnow data warehouse. This was then used to focus marketing on this banks, reducing marketing expenses by 80%.

Technology: Python, SQL, KNN, Kmeans
Project: ETL threading optimization using machine learning

• In order to refresh the data warehouse nightly, 3000 tables are retrieved from the bank and then inserted into the data warehouse. Unfortunately, random errors would occur at irregular intervals causing a failure in

the process. The errors in question had been occurring for 10 years. I used several mathematical techniques to model the ETL process. The model and subsequent analysis uncovered 2 bugs in the underlying Microsoft Framework. The workaround reduced errors by 90% and increased throughput by 50%. The approach used is similar to financial portfolio analysis.

Technology: Python, SQL (Mathematics (Correlation Matrix, ANOVA)

Senior Software Engineer

PNC - Overland Park, KS - January 2006 to March 2013

Responsibilities:

• Developed commercial mortgage software using VB.NET, SQL and PDF.

Previous experience available upon request

EDUCATION

Master of Science in Business Intelligence and Data Analytics

Rockhurst University - Kansas City, MO

Master of Science in Human Performance Technology

Boise State University - Boise, ID

Bachelor of Science in Applied Engineering and Technology

Thomas Edison State College - Trenton, NJ

CERTIFICATIONS/LICENSES

Senior Reactor Operator

May 1989 to May 1995