MATT PETTIS

I have acquired deep experience with data science, data engineering, and model building over the span of my career. I have also gained a breadth of experience applying this over such sectors as the legal field, pure IT, retail software, and the industrial/manufacturing sector. My expertise has become identifying the heart of the problem, creating helpful models and analytics, and being able to work from existing legacy systems to put into place actual functioning models into production. I am passionate about bringing models from ideas and experiments into everyday business production workflows.

🖥 WORK EXPERIENCE

2019 | Current

Senior Data Scientist
Trane Technologies   📍 White Bear Lake, MN

- Oversee existing machine learning projects with third parties. Assess data science work of these parties, supply data, critique approaches and results, recommend solution purchases when appropriate.
- Create models combining weather, building occupation, and HVAC system performance in order to optimize HVAC running policies. Designing the deployment strategy for the model implementation and lifecycle.
- Create instrumentation tooling around new and extisting system APIs. Created software to extract and display process DAG-dependencies. Create batch processes from ReST-ful APIs. Created software to translate API JSON payloads to editable CSV files and back to JSON format for our QA team so they could easily create functional tests.
- Create automation tooling through the AWS CLI to automate Cloudwatch querying and analysis with offline tools.
- Created automation tooling around gitlab to automate reporting around codebase commit and promotion.
- Legacy software system design and migration. Design and evaluate system migration between ReSTful-centric dataflow model to GraphQL-centric model.
- Manage critical software team that is writing/rewriting R code to do custom analytics for flagship HVAC control and alerting offering (TraneConnect), as well as code and provide architectural frameworks and processes. Highlight: perfect yearly performance review for my efforts leading this team for 2021.
- Consult on architecture design for data lake corporate data strategies. Consult on existing domain data modeling to handle structure and time-dependent requirements.

CONTACT INFO

✉ matthew.pettis@gmail.com

⌨ github.com/mpettis

📞 +1 612-232-9212

in matt-pettis-79388b4

For more information, please contact me via email.

SKILLS

Areas of Expertise: Machine Learning (Classical models and ANN/Keras), ETL and Data Engineering, ReSTful APIs.

Have led teams, managed people, and been an individual contributor.

Time Series Analysis

SQL Expertise

Command-line Expertise, Linux: Bash tools, piping, batch processing.

Programming: Advanced - R, Python; Formerly Advanced - Perl, SAS; Novice/Hobby: Clojure, Scala.

Worked mostly in AWS stack, some Azure.

Principal Data Scientist

Honeywell   📍 Golden Valley, MN

• Manager of data science team. Directed work of data scientists, managed projects for my reports, gave annual reviews, mentored my reports on data science, coding, and career path questions.

• Lead on productization of ML models for client. Full-stack developer and manager for project. Work includes

  • Project managing creation and deploy of multiple models. Managing work through sprints in Atlassian JIRA. Run stand-up meetings for the project.

  • Managing model assessment and selection with the junior data scientist developing the models.

  • Administering the Bitbucket git repos for the project.

  • Coding the models, collaborating with model creators on such. Enforcing code standards for all code contributors. Creating tests and run harnesses for code. Doing CI via Bamboo.

  • System administration of Microsoft R Machine Learning system, which hosts our models online.

  • Architecting the interfaces (defining JSON interface and contracts).

  • Architecting the interaction of our models with the legacy systems.

• Created and enforced data science practices and workflow structures. To that end, created and administered git repositories that enforced standard workflows and processes, taught and enforced with peer teams.

• Sales forecasting for catalyst divisions. Implemented time series models (ARIMA, Holt-Winters, STL, linear regressions) in R. Administered Nutonian/Eureqa toolset for fellow data scientist to test models on it. Presented to SVP level management on findings. Managed data flow of external features, which had dozens of sources (used R + make technology to manage acquisition of related features, which numbers in the thousands).

• Coordinated team of people to evaluate machine learning models against manufacturing plant data (airplane brakes, data types POMSnet and Uniformance.) Both the automation of simple model coverage and more powerful models and techniques (SVM, neural nets, Quadratic Linear Discriminants for models, XGBoost and ensemble models for techniques) were employed.

• Led groups of data scientists and developed processes with which multiple data scientists could create models and assess them against one another. Done via standardizing model outputs that were comparable across models.

• Worked with business partners to define scopes of machine learning processes and desired outcomes, as well as other desired deliverables (such as general analytic capabilities).

2016
|
2018

Contractor

Best Buy   📍 Richfield, MN

- Part of team to implement a data lake for Adobe Web analytic data. Worked with data ingest, field validation, find and fix match-merge quality problems with data.
- Created regression models to assess impact of online product rating systems on purchase lifts.
- Performed time-series modeling to predict web traffic composition by platform.
- Trained analysts in using R, SQL, and Hadoop to automate recurrent jobs.

Senior Data Scientist

Quantum Retail   📍 Minneapolis, MN

Senior Scientist, responsible for client configurations of Quantum's software, creating generalized models of seasonal time series indexes for generic re-use by the company.

- Analyze client data, translate into data models and software configurations for the client, and work with the client to explain decisions as well as integrate their business intelligence.
- Lead teams of other scientists for client work as warranted.
- Utilized k-means and hierarchical clustering to identify client attributes that have the best predictive power for our forecasting system.
- Developed jobs and data models on Hadoop Hive to transfer existing processes that were run on Oracle to be run on Amazon Web Services.
- Create a framework of repeatable processes of analysis for other scientists and analysts to follow. Adopted R-language 'knitr' for a reproducible analysis framework, created process to quickly deploy the framework (and webserver) on our infrastructure. Trained others in how to use knitr, and how to collaborate on these scripts via git.
- Rewrote company's core seasonality index generation, which is the backbone of Quantum's forecasting system off of client data, in collaboration with one other scientist in 6 weeks. Previous rewrite took 1 year. Focus was not only on the algorithm, but instrumentation of the process so that edge cases were easily fixable by client, and made extensible to accommodate future unknown algorithms.
- Designed a more flexible software data model for our core product to leverage the seasonal indexes generated by our rewritten seasonality process.
- Ran in-house professional development talks on how to use core technologies to accomplish company's data science needs in an efficient manner.
- Conceived of and coded deployment health checks that were previously non-existent and made into reusable, shareable code for all scientists to use.

1999
|
2012

Lead Systems Analyst
Thomson Reuters   ◉ Eagan, MN

Lead Systems Analyst, responsible for capacity planning, business forecasts of time series data, created diagnostic scripts for first-level support workers.

- Responsible for creating forecasts of usage and the hardware capacity plans for the flagship Westlaw application, translating into capacity plans on the order of magnitude of millions of dollars.
- Was company expert in SAS, both in coding and software configuration and deploy. Utilized SAS/ARIMA models for the time series forecasts.
- Championed change of forecasting model from one based on pure accuracy to one that incorporates risk analysis and impact of under/over forecasting. This led to better decision making on purchases and understanding of risks taken.
- Designed and coded department-wide software that monitored various health aspects of our software and hardware systems.
- Designed and coded department-wide software that allowed for automated health checks of various systems, coding up the top 10 most common problems experienced by the support teams and automating their diagnostic information collection. This drastically reduced the response time of our support teams in

🎓 EDUCATION

1997

Northwestern University
M.S. in Environmental Engineering   ◉ Evanston, IL

1995

Northwestern University
M.S. in Mathematics   ◉ Evanston, IL

1994

Gustavus Adolphus College
M.S. in Physics, M. A. in Mathematics   ◉ St. Peter, MN

👤 TEACHING

2021
|
Present

Programming Logic and Design, in Python
Minneapolis Community Technical College   ◉ Minneapolis, MN
Teaching course on the basics of computer programming via Python.

2020

Time Series Forecasting, in R (MSBA 8160)
Hamline University   ◉ St. Paul, MN
Teaching course on the basics of time series forecasting in R.

📄 OTHER NOTES

- **Code examples**

  - Time zone manipulation in R, delivered to R User Group: https://gist.github.com/mpettis/1ae48bd46e77e67bccd7fb43a8ba2a8d
  - dplyr and plyr package, delivered to R User Group: https://gist.github.com/mpettis/70dcb33f7328e21ec485fdf8727c97ef
  - Non-standard evaluation and standard evaluation in R, delivered to R User Group: https://gist.github.com/mpettis/5fc5e2d2d7d54291e047e59fb5baf1ea
  - Reinforcement Learning for simple thermostat, submitted for use for documentation with tensorflow package: https://github.com/tensorforce/tensorforce/pull/671