Matthew Evan Taruno

DATA SCIENTIST
mtaruno@u.rochester.edu • www.matthewtaruno.com

EDUCATION

University of Rochester | Bachelors of Science in Data Science, Minor in Business

MAY 2017 – 2021, ROCHESTER, NY | GPA: 3.3

- **Extracurriculars:** Pledgemaster at Alpha Kappa Psi, Men's Club Basketball, Piano at Eastman
- **Relevant Courses:** Data Science at Scale, Predictive Analytics for Business, Database Systems, Data Structures and Algorithms, Statistical Speech and Language Processing, Bayesian Inference, Corporate Finance, Operations Research

BrainStation | Diploma Program, Data Science

MAY 2020 - AUGUST 2020, ONTARIO, CA

SKILLS

- **Languages: Python, R, Java, and SQL**
- **Tools and Frameworks:** Tableau, git, CLI, scikit-learn, pytorch, tensorflow, keras, pyspark, databricks, pandas, numpy, seaborn, plotly, TS fresh, spaCy, delta lake, MLflow, gensim, and excel
- Proficient in **data analysis**, **machine learning, bayesian statistics**, and **natural language processing** for multiple applications including conversational interfaces, financial markets, customer churn, and sensor data
- Native fluency in English and Indonesian, workplace level proficiency in Mandarin
- Interests: Music production, chess, stock market, and table tennis

PROJECTS

Research Project: Data Literacy Expectations   JAN 2021 - Present

- Collaborating with Sandra Cannon to create a useful metric for data literacy expectations for job recruiters using text mining, topic modelling, and supervised learning techniques based on a corpus scraped from LinkedIn using Selenium

Vnomics DPF Failure Prediction   JAN-MAY 2021

- Successfully built autoencoder models with ML Flow and Keras to predict truck failures given sensor data for a fuel optimization startup called Vnomics as my capstone project; the model is optimized by comprehensive time series feature engineering with TS Fresh to achieve a high recall score of over 80% on unseen data

EVE Bot   AUGUST 2020

- Created a Twitter Apple Support chatbot leveraging NLP techniques such as named entity recognition and document embeddings on Twitter data and a Bidirectional LSTM in Python to classify intents and deployed using Streamlit
- Wrote a Medium article that contains a video demo and goes in depth on the technical workings of this project

EXPERIENCE

Machine Learning Project Lead | STEM-Away   SANTA CLARA, CA; JUNE 2021 - Present

- Leading a team in collaboratively replicating and extending the research of Percha and Altman 2015 "Learning the Structure of Biomedical Relationships from Unstructured Text"
- Leveraged an XML parser for ETL from Medline and PharmGKB drug bases, Stanford parser Java implementations to create dependency matrices, and spaCy for dependency parsing for extracting useful lexicons and semantic signals

Technical Project Manager | Valdo Incorporated   JAKARTA, INDONESIA; JAN 2019 - Present

- Automated the Citibank credit card application process using OCR/ICR technology and machine learning methods for document template classification to reduce manual labor and increase precision in collaboration with Laiye
- Implemented a new ID card image to tabular data entry system with OpenCV and BotScript for the computer vision implementation to improve ID processing speeds for TokoPandai