mindyng88552@gmail.com
mindyng.github.io
5105082455
www.linkedin.com/in/mindyng85
mindyng

I am passionate about combining descriptive analytics with results-oriented data problem-solving and bridging the knowledge gap across multiple disciplines and presenting insights/results to different audiences and teams.

Skills

PROJECT MANAGEMENT
Scoping out Business Problem
Defining Project Success
Metrics Development
Defining KPI's
Team-Player
Cross-Discipline Collaboration
Insights to Stakeholders

LANGUAGES
SQL
Python

DATA ENGINEERING (ELT)
Dataform
dbt
PostgreSQL
Meltano
Snowflake
Google Cloud Platform

DATA WRANGLING
Data Cleaning
Data Integrity Checks / Assertions

STATISTICS
Descriptive Statistics
Inferential Analytics
Hypothesis Testing
A/B Testing

MODELS / MACHINE LEARNING
Linear Regression
Logistic Regression
Natural Language Processing (NLP)

BUSINESS ANALYTICS
Time Series Analysis
Churn Prediction

VISUALIZATION / BUSINESS INTELLIGENCE
Looker
Superset
Tableau

MINDY NG
DATA SCIENTIST

Employment

Mercari US — Palo Alto, CA
Data Scientist — Sept. 2022 to Oct. 2022
- Helped build out search team's new metrics table in Dataform to power dashboards and assess experiment results.
- Analyzed price elasticity between highest GMV/most searched for category groups to understand buyers' demand change in relation to item price changes.

Mercari US — Palo Alto, CA
Business Intelligence Analyst — Aug. 2021 to Aug. 2022
- Created Search RFM segmentation to personalize search and increase GMV.
- Investigated query chaining to understand search engine performance as well as searchers' persistence for casual and power users' high value purchases.
- Performed user journey analysis across web to provide insights on platform's highest touch points for ML efforts to converge on.
- Analyzed navigation vs search activity to improve our UI to cover all our users' intent to drive north star metrics.
- Built statistical significance framework for experimentation at scale to add criteria for product roll out.

Forethought — San Francisco, CA
Implementation Engineer — July 2020 to Sept. 2020
On the Customer Experience team, leading all technical requirements and touching all aspects of the business: Engineering, Product, Sales and Customer Success
Implemented: State-of-the-art NLP models to help clients be geniuses at their job
Involved: Data Engineering, Data Science, Machine Learning/Artificial Intelligence, Business Intelligence -- owning whole data pipeline Post-Sale

- Queried MongoDB to create customer business rules.
- Designed AI Training datasets to feed into XLNet and BERT models using Jupyter Python notebooks.
- Analyzed trained models' performance to deploy best automated NLU models for clients.
- Verified live models' predictions were successful via API calls to clients' Salesforce Help Desks.
- Reduced client's SPAM from 64% to less than 1%.
- Helped save client >$20,000 in human labor cost from Customer Support Agents manually labeling tickets.
- Completed data analysis that contributed to signing of >$400,000 deal with Instacart.

Immuno Concepts — Sacramento, CA
Quality Control Analyst — July 2010 to Apr. 2019
- Built linear regression models to determine whether or not products were drifting from quality.
- Tracked trends and outliers to make manufacturing recommendations to management to create efficiencies and increase profit margins.
- Created product performance reports to drive key business investments for following quarter.

University of California, Davis — Davis, CA
Research Associate — Jan. 2005 to Dec. 2008
- Through repeated experimentation explored sigma70 subunit architecture to characterize macromolecular complexes involved in transcription of growth-related genes.
- Narrowed down which protein chain substitution in antibody-derived proteins fit best with research aims in pre-targeting radioimmunotherapy for Non-Hodgkin's Lymphoma.

Projects

Deployed Web App for Business Stakeholder — June 2021 to June 2021
Took a business stakeholder's question, translated it into a data question, performed ETL and created a web app with visualizations and analytic conclusion. Deployed web app with non-technical language using Streamlit's Sharing feature.

Modern Analytics Data Stack Built from Scratch — Feb. 2021 to Feb. 2021
Built from the ground up and maintaining- data pipeline to perform ELT and BI visualization layer. Stack included open source: PostgreSQL DB, Meltano, dbt and Superset.

Time Series Forecasting on Uber Eats' Vendors — Dec. 2018 to Dec. 2018
Utilized 7,911 samples of date-stamped data and predicted which vendors were worth continuing business with based on ROI.

Trended each vendors' data with Facebook's Prophet. Trends performed over a span of 15 months. Data further broken down into weekly and daily trends. Resulting model performance based on 30-day horizon producing 0.01 - 0.03 RMSE.

Postmates New Market Analysis with Geospatial Heatmaps — Mar. 2019 to Mar. 2019
Analyzed 3-sided market to explore contributors to conversion and churn, used heatmaps to visualize supply and demand, determined health of market and addressed data integrity issues.

Medicare Prescription Drugs Analysis — July 2019 to July 2019
Analyzed 25,209,130 samples of Medicare Part D Prescription use to determine how geography correlates with provider density, provider specialties and drug costs.

Plotly and Seaborn used to visualize number of providers across states, to geocode provider specialties and to examine differing degrees of drug cost variance across the U.S.

Publication

American Chemical Society Publications · Mapping Protein–Protein Interactions by Localized Oxidation: Consequences of the Reach of Hydroxyl Radical Feb. 2009 to Apr. 2009
Results provided fundamental information for interpreting protein foot-printing experiments in other systems.

Volunteering

CoronaWhy — Apr. 2020 to June 2020
Machine Learning Engineer
Helping to fight against Coronavirus.

CoronaWhy is a globally distributed, volunteer-powered research organisation of 1000+ members. We're using DS and AI to assist the medical community and policy makers answer key questions related to COVID-19. It's supported by Google, Amazon, NASA and other companies.

I am embedded within the Vaccine/Therapeutics Task team, helping the Paper Study Classification group build baseline models to filter papers based on study design.

Education

Springboard, Data Science Career Track — Jan. 2017 to Dec. 2017

University of California, Davis — Sept. 2003 to Dec. 2007
Genetics Bachelor's of Science