NATALIE CHEN

DATA SCIENTIST

CONTACT

- ✉ chennat811@gmail.com
- ☎ 650-892-6607
- ⚲ Santa Clara, CA 95051
- in linkedin.com/in/Natalie-Chen8
- ⌂ chennat811

SKILLS

PROGRAMMING: Git, SQL, Python

MACHINE LEARNING:
Linear/Logistic Regression,
Tree-based methods,
Ensemble methods,
Natural Language Processing, Keras

LIBRARIES: Pandas, Numpy,
BeautifulSoup, NLTK, Gensim,
Statsmodel, Scikit-learn, Seaborn

**GEOGRAPHIC INFORMATION
SYSTEM:**
Supervised classification,
Change detection,
Time Series Analysis

EDUCATION

Harvard University
Sept. 2017 to Current
Master's of Liberal Arts,
in Extension Studies:
Sustainability
Working Thesis Title:
Regenerative agricultural
practices in California's
Central Valley for soil health
and farmer profitability.

University of
California, San Diego
Sept. 2012 to Sept. 2015
Bachelor's of Science
Molecular Biology

PROJECTS

Plant Disease Detector using Deep Neural Networks
- Applied transfer learning with DenseNet169 to train and retrain a neural network for the PlantVillage dataset.
- Obtained a test accuracy of 98.7% and macro average recall and precision values of 0.98.
- Created a Streamlit app thats deployed on Heroku. It takes an user input and provides a prediction, summary of the disease, and management recommendations scraped from Wikipedia. https://plant-disease-detector-88.herokuapp.com/

What did you say, clean meat?
- Scraped 100,000 Reddit posts and comments using the Reddit API for the keywords "lab-grown meat" and "clean meat".
- Utilized NLP techniques such as latent dirichlet allocation to determine the topics that the Reddit community was interested in.
- Analyzed and compared the sentiment of each topic and found that "lab-grown" is slightly more negative than "clean meat" for certain topics.

Soil Erosion Tolerance Predictor
- Used Tableau for exploratory data analysis.
- Built a Random Forest Classifier using six features with a 95% accuracy in predicting whether soil erosion is tolerated.
- Created a Streamlit app that utilize user input to predict soil erosion tolerance.
- Utilized QGIS to map the predicted vs actual soil erosion tolerance.

IMDB User-Rating Predictor
- Scraped IMDB and Box Office Mojo data with BeautifulSoup.
- Built a linear regression model with engineered features, to predict the user ratings of documentaries and has an MAE of 0.35.

Forecasting Vegetation Recovery after a Wildfire
- Final project in Harvard's ISMT-E-158 - Remote sensing data and applications.
- Utilized QGIS to look at change difference in vegetation, moisture, land surface temperature, and burn index of Trinity Alps, CA.
- Created a time series of derived indices from Landsat images from 2008-2020.

EXPERIENCE

Metis
San Francisco, CA
Data Scientist
Sept. 2020 to Dec. 2020
- Completed an immersive 12-week accredited data science bootcamp with a strong focus on project-work. Designed, completed, and presented the projects. *See projects section.*

Indigo Agriculture
Boston, MA
Lab and Facilities Operations Technician
July 2019 to Mar. 2020
- Maintained safe laboratory operations, produced media for the Scientists, purchased supplies/equipment and evaluated/created SOPs.
- Arranged sustainability initiatives to promoted environmentally friendly practices in the workplace.
- Organized, hosted and presented in sustainability-related best practices and recent events.

Poseida Therapeutics
San Diego, CA
Research Associate
Dec. 2016 to Mar. 2019
- Independently designed, managed and supervised the Gene Therapy department's cloning projects for proprietary Piggybac transposon and gene-editing technologies.
- Managed and conducted multiple types of experiments, such as bone marrow analysis, vector copy number, and protein expression assays.
- Analyzed data from each experiment and present data at company meetings with active participation in strategic decision-making.