

# JESSIE LIN

☎ (530) 760-9746 · ✉ [jessiejingwen@gmail.com](mailto:jessiejingwen@gmail.com) · 🌐 [jessie-jingwen-lin.github.io](https://jessie-jingwen-lin.github.io) · 📄 [github.com/Jessie-Jingwen-Lin](https://github.com/Jessie-Jingwen-Lin) · 🔗 [linkedin.com/in/jiwlin](https://linkedin.com/in/jiwlin)

## SKILLS

---

Python (Pandas, Numpy, Matplotlib, Scikit-learn), SQL (MySQL), Git, Shell, R (ggplot, dplyr), Data Scraping, Modeling (regressions, classification, clustering), Data Visualization (Python, R)

## PROJECTS

---

### Network Clustering Analysis and Visualization of Popular Twitter Hashtags Sep 2021

- **Scraped and cleaned** a dataset of 100,000+ tweets, and used Python to insert the data into a **MySQL database**.
- Computed hashtag co-occurrence rates via **SQL queries** on the database.
- Performed **graph clustering and layout** on this graph of hashtag co-occurrences to identify and visualize different hashtag communities.

### Kaggle Challenge: Analytics and Predictive Modeling of HR Data July 2021

- Conducted **exploratory data analysis** to understand the data and visualize variable correlations.
- Preprocessed data by performing **data imputation** for missing values and **data augmentation** to address imbalanced classes.
- Trained a **Catboost classifier** to predict if an employee leaves their job, and tuned hyperparameters using **k-fold Bayesian optimization**.
- Evaluated model performance on test set via **accuracy and ROC curve** with the AUC difference  $< 1.5\%$  between the train and test sets.

### Stock Data Analytics for Overnight Trading May 2021

- **Scraped and cleaned** stock price data for 10,000+ stocks across 800+ days from NASDAQ, NYSE and Yahoo Finance.
- Ranked tickers by computing statistical metrics over multiple time windows using **Python and Pandas**.
- Presented the results in an interactive web app built with **Flask and jQuery**.

## RELEVANT COURSES

---

Introduction to Data Analysis in R at UMass Amherst, Machine Learning A-Z (Udemy), Supervised Learning in R: Classification (DataCamp), Time Series Analysis in R (DataCamp), Probability and Mathematical Statistics (A), Advanced Mathematics (A), Fundamentals of Program Design (A), Linear Algebra (B)

## EDUCATION

---

**University of Massachusetts Amherst**, *Master in Food Science*, GPA: 4.00 May 2020

**Isenberg School of Management – UMass Amherst**, *Intensive Business Training Program* Jun 2019 - Aug 2019

**University of California Davis**, *Study Abroad*, GPA: 3.78 Sep 2016 - Jun 2018

## EXPERIENCE

---

**American River Nutrition LLC**, *Lab Technician, Intern* May 2019 - March 2020

- Set up HPLC method and conducted product quality assurance using HPLC and GC.
- Reduced material cost ~80% by developing new extraction methods for an alternative raw material.

**Colon Inflammation and Cancer Lab, UMass Amherst**, *Graduate Researcher* Sep 2018 - Sep 2020

- Performed mice study on the effect of jackfruit extract on colorectal cancer development.
- Screened 45 different fruits and vegetables on the inhibition effect of cytochrome P450 enzymes.

**Gut Microbiome Lab, UC Davis**, *Undergraduate Researcher* Jan 2017 - May 2018

- Analyzed bifidobacteria DNA sequences using Basic Local Alignment Search Tool (BLAST).
- Increased plasmid transformation efficiency of bifidobacteria by 80% using electroporation.

PUBLICATIONS: [SEE GOOGLE SCHOLAR](#)

---