

YEWEN ZHOU

<https://www.linkedin.com/in/yewen-zhou/>
<https://hegelim.github.io/personalwebsite/>

(626) 492-9028
yz4175@columbia.edu

EDUCATION

Columbia University M.S. in Data Science <ul style="list-style-type: none">GPA: 3.83 / 4.0Coursework: Algorithms, Big Data, Causal Inference, Machine Learning, Applied Deep Learning, Finance for DS	New York, NY Dec 2022
University of California, Berkeley B.A. in Data Science, Business Analytics Concentration <ul style="list-style-type: none">GPA: 3.93 / 4.0, Phi Beta Kappa SocietyCoursework: Data Structures, Time Series, Artificial Intelligence, Probability and Statistics, Decision Analytics, Intro to Finance	Berkeley, CA May 2021

SKILLS & TECHNOLOGIES

Programming:	Python, Jupyter, Linux, SQL, R, Java, HTML5, CSS, JavaScript
Python Packages:	pandas, pyspark, numpy, scipy, scikit-learn, matplotlib, seaborn, re
Frontend Frameworks & Cloud Services:	Django, Bootstrap, Plotly, Google Cloud Platform
Writeup:	Markdown, reStructuredText, LaTeX

WORK EXPERIENCE

SAFE Lab, Columbia University Data Scientist <ul style="list-style-type: none">Matched 200 medical notes based on cosine similarities; trained logistic regression classifier on Bag of Words (BOW) and TF-IDF matrices with hyper-parameter search, achieving cross-validation recall 0.99Combined 3 tables with more than 2,000 rows and grouped with datetime intervals for each medical record number (MRN), allowing convenient table lookup for team members	New York, NY Oct 2021 - Present
iQIYI, Inc. Ads Algorithm Backend Intern <ul style="list-style-type: none">Developed a testing framework for ads allocation emulator with more than 10,000 records; deployed in the server launched overseas in more than 5 countriesCreated a SARIMA time series module for ads inventory prediction, achieving a cross-validation accuracy of over 80%Implemented High Water Mark (HWM) algorithm from scratch leveraging logging, numpy, pandas based on Yahoo research paper for compact allocation; used as the 1st version by algorithm and product teams of more than 10 people	Beijing, CN May 2021 - Aug 2021
Goodly Labs, Inc. Validation Team Lead <ul style="list-style-type: none">Led a team of 3 undergraduate students supporting Public Editor validation with more than 1,000 input entriesLaunched a program based on Krippendorff's alpha on Jupyter Notebook utilizing pandas, numpy, providing insights into level of agreement among users	Berkeley, CA Aug 2020 - May 2021

PROJECTS

Columbia University, Realtime Twitter Sentiment Analysis <ul style="list-style-type: none">Developed 6 ML models including Linear Regression, Ridge Regression, Gradient Boosting, AdaBoost, Random Forest, and SVR for aggregated twitter sentiment prediction, attaining test RMSEs less than 0.1Leveraged Virtual Machine (VM) on Google Cloud Platform (GCP) to decrease model training time by 16xCreated a dashboard using Bootstrap, Django, HTML5/CSS/JavaScript, and Plotly, displaying real-time Twitter sentiment prediction	Nov 2021 - Dec 2021
Columbia University, Stock Price Prediction <ul style="list-style-type: none">Utilized Airflow Scheduler to collect stock prices from 5 tech companies automatically daily at 7 amTrained and updated 5 linear regression models for stock price prediction; obtained relative errors less than 0.01	Nov 2021 - Dec 2021
Open Source, The solveminx Python Package <ul style="list-style-type: none">Implemented an object-oriented, open-source Python module to solve a sum of min and max equations applying regular expressions, numpy, sympy, and matplotlibDesigned unit tests using pytest to validate module extensively with more than 30 testing casesDistributed on the Python Package Index (PyPI) with documentation hosted on GitHub written in reStructuredText and Markdown	Jul 2021 - Sep 2021