LUCY WANG

wangkehua@berkeley.edu | (510)-822-1176 | 2024 Durant Ave, Berkeley, CA 94704 | www.linkedin.com/in/lucy1116

EDUCATION

University of California, Berkeley | Berkeley, CA **Expected Graduation: May 2021**

B.A. Computer Science, B.A. Statistics(GPA: 3.70)

- Relevant Coursework: Machine Learning, Artificial Intelligence, Natural Language Processing, Computer Vision & Photography, Optimization Models, Data Science Techniques, Algorithms, Databases, Operating Systems
- Awards: Individual 6th place, North Carolina State Mathematics Contest Finals; Dean's Honor List

PROFESSIONAL EXPERIENCE

SurveyMonkey | San Mateo, CA **May 2020 – August 2020**

Software Engineer Intern

- Developed Deep Flakes project which uses ML algorithms to predict flaky tests during automation testing. It has two stages:
  - Built Data Extraction API to transfer data from Jenkins to MySQL database using Pyramid Framework and REST API.
  - Created a dashboard with React and JavaScript for engineer users to view the flaky test prediction.
- The Deep Flakes project is used in production, which shortened engineering testing cycles and debugging time.

Chinese BookNLP Research | Berkeley, CA **June 2020 – Present**

Research Assistant with Professor David Bamman

- Add Chinese linguistic features to BookNLP toolkit, which is used to analyze linguistic structure of text in literature.
  - Performed NER and Coreference Resolution annotation and data processing on 20+ Chinese literature books
  - Trained NER model with biLSTM on annotated corpus and character/word level embedding. Achieved 75% accuracy.
- Next Step: Implement BERT to improve accuracy for NER, Coreference Resolution, and syntactic structure prediction.

Orbis AI | Berkeley SkyDeck, CA **August 2019 – May 2020**

Machine Learning Intern

- Trained and tested CNN-based text-to-speech models for synthetic voice generations with Tacotron2 and Pytorch. Performed hyperparameter tuning with voice decoders Waveglow to achieve better pronunciation of the voice.
- Models are currently deployed for services, which generated $50k revenue for the company.
- Explored designing Automated Voice Model Testing Tool by converting generated audio to text through Google Cloud Platform, and use BLEU to estimate similarity to the original text.

Smartdot Technology | Beijing, China **May 2019–August 2019**

Software Engineering Intern

- Developed rule-based algorithms for NLP user-defined NER in the legal domain, which improved the accuracy of our smart contract review platform by 15%. Annotated texts for NLP model training.
- Developed risk analysis feature for the platform based on review rules to alert users about potential risks in the contract, which generated $30k in revenue for the company.
- Technologies: HanLP(NLP tool package for the Chinese language), Python, Django, Vue, MySQL

HasBrain | San Francisco, CA **November 2018 – May 2019**

Data Analyst Intern

- Generate contents for HasBrain's data science educational platform by analyzing open-source data sets using Python. Steps include exploratory analysis, visualization, and train model with linear regression and classification.
- Examples of contents: applied linear regression techniques to predict bike rental usage in DC to 92% accuracy.

SELECTED PROJECTS

Spam/Ham Classification with Random Forest **March 2020**

- Trained decision trees and random forest models with bootstrapping and cross-validation from scratch for Spam/Ham classification, which achieved 85% prediction accuracy in Kaggle. Investigated in early-stopping criterions.

Predicting Taxi Ride Duration **May 2019**

- Applied EDA and PCA to assess the impact of historical events on taxi ride durations and filtered the data accordingly.
- Trained model using different methods including linear regression and decision tree regression and compared results.

SKILLS & INTERESTS

Technical Skills: Python, SQL, PyTorch, Numpy, Keras, scikit-learn, Nvidia server, AWS, Google Cloud, Java, C, JavaScript
Interests: Swimming, Badminton, Nature, 6-year piano experience