



# Yinkai Wang

18650146958 | ywang88@gmu.edu  
yinkaiw.github.io

## EDUCATION

### George Mason University

Computer Science Bachelor VSE bachelor

- Honors/Awards: Dean's List(2018-2020)

Aug 2018 - Aug 2021

Fairfax

### Huaqiao University

Computer Science Bachelor VSE bachelor

Aug 2017 - Aug 2022

Xiamen

## PUBLICATIONS

### Ensemble Machine Learning System for Student Academic Performance Prediction (Accepted by Workshop for Undergraduates in EDM at the Fourteenth International Conference on Educational Data Mining)

Joint first author

- Used ensemble machine learning system to predict students' final grade.
- Due to the influence by COVID-19, the education to students faced to a severe problem: It's much more harder for teacher to help student and know students' learning condition.
- This model consists of two components, ensemble feature engineering module and ensemble prediction module. Extensive experiment results have shown the superiority of our model over other traditional machine learning models, both in stability, efficiency and accuracy.

### Deep Latent-Variable Models for Controllable Molecule Generation (BIBM 2021 in submission)

The second author

- Proposed several deep latent-variable models to generate small molecules with desired molecular properties.
- The models operate under supervised, disentangled representation learning and leverage both graph representation learning to learn inherent constraints in the chemical space and inductive bias to connect chemical and biological space.
- The evaluations show that the models are a promising step in controllable molecule generation in support of cheminformatics, drug discovery, and other application settings.

## RESEARCH EXPERIENCE

### Multilingual Geospatial Language Expression Discovery

Research Assistant(Advisor: Prof. Antonios Anastasopoulos)

- Take parallel data between English and another language, Use existing tools to discover GLEs in English.
- Align the parallel English-X data, Project the labels from English to X.
- Compare the projected annotations on cmn, ell, ita with the ones produced by the in-language models.

### Diffusion Probabilistic Models for Protein Generation

Research Assistant(Advisor: Dr. Amarda Shehu)

- Use diffusion probabilistic models to generate protein.
- Use short range and long range to evaluate the result.

### Predicting Minimum Inhibitory Concentration for Quaternary Ammonium Compounds w/ Machine Learning

Research Assistant(Advisor: Dr. Amarda Shehu)

- Create three settings based on the 70 features of ~450 Quaternary Ammonium Compounds.
- Predict four Minimum Inhibitory Concentration values with ten regression machine learning models.

## PROFESSIONAL EXPERIENCE

### Bytedance

Intern DevEco

Apr 2021 - Jul 2021

Beijing

- Focus on the base of host app of android, which have the coupling relationship with most of app from bytedance, like tiktok, lark, Toutiao...
- Create a mock setting environment implement to help QA test, greatly improve the efficiency of testing and publishing.
- Handle unexpected technology problems of user. Judge the urgency of problems and solve it.
- Make a great and comfortable environment for all the developer who are developing microapp on bytedance.

### Google Kaggle

Google Smartphone Decimeter Challenge

- Ranked top 20% on the Google Smartphone Decimeter Challenge w/ a group of five undergraduate students.
- Designed data cleaning, preprocessing, data analysis, model selection, result evaluation and visualization pipeline.
- Mastered real-world data science challenge with machine learning pipeline and team collaboration.

## SKILLS LIST

**Programming skills:** Python, Java, C, Kotlin, MIPS, Julia

**Research Interests:** Machine Learning, Deep Learning, AI for Science, Deep Graph Learning, Natural Language Processing