

# Vedang Waradpande

30 Newport Pkwy, Apt 2712, Jersey City, NJ 07310

+1 (848) 437-1822

[vedang.waradpande@gmail.com](mailto:vedang.waradpande@gmail.com)

[vedangw.github.io](https://github.com/vedangw)

[LinkedIn](#)

[GitHub](#)

---

## EDUCATION

- M.S. in Computer Science - Rutgers University, New Brunswick** (GPA: 3.92/4.0) **Sep 2021 – May 2023**  
*Concentration: Machine Learning*
- B.E. (Honors) in Computer Science - Birla Institute of Technology and Science, Pilani** (GPA: 7.82/10.0) **Aug 2015 – Jul 2019**
- 

## RESEARCH EXPERIENCE

**Research Teaching Specialist III** **New Jersey Medical School, [Rutgers University](#)**  
*Newark, NJ, USA* **Oct 2022 – Present**

- Advised by Dr. Joel Freundlich, collaborated with Dr. Connor Coley (MIT). Area of research: Machine Learning, Drug Discovery.
- Developed ML models for efficient search of drug candidates for Tuberculosis through molecular property prediction.
  - Predicted the [PK profile](#) of molecules by training tree-based models (XGBoost, LightGBM, CatBoost, and Random Forest) & fine-tuning state-of-the-art deep ML models (Graph Neural Networks, Transformer-based) with a test AUROC of **93%**.
  - Developed hand-crafted heuristics for aggregating results of several models and selected 96 compounds for experimental validation, resulting in a **43%** PK activity rate and several candidates potentially moving to further testing stages.
  - Working on a preprint towards the *Nature Machine Intelligence* journal.
- Working on predicting docking ability of ligands with KasA protein for TB using deterministic & ML methods (e.g. DiffDock).
- Contributed majorly to developing the lab's ML Python library and setting best software engineering practices.

**Student Researcher** **Computational Biomedicine Imaging and Modeling Center, [Rutgers University](#)**  
*New Brunswick, NJ, USA* **Sep 2022 – Present**

- Advised by Dr. Matthew Stone and Dr. Karl Stratos. Area of research: Natural Language Processing.
- Analyzed the ability of language models such as T5-lm-adapt, T0, GPT-Neo, and GPT-3.5 to understand and complete analogies from a philosophical and neuroscience perspective.
  - Achieved a 20 point increase in MRR over baseline with efficient prompt engineering and In-Context learning.
  - Working on a preprint towards the *ACL 2024* conference.
- Adapted the [UPR method](#) using the T0 language model towards Question Answering, Fact Checking, and Entity Linking. Achieved competitive top-100 accuracy on Meta AI's KILT datasets such as TriviaQA (**86.5%**), NQ (**85.15%**), FEVER (**58.2%**).

**Research Assistant** **Department of Genetics, [Rutgers University](#)**  
*New Brunswick, NJ, USA* **Jan 2022 – May 2022**

- Advised by Dr. Premal Shah. Area of research: Machine Learning, Computational Genomics.
- Developed ML models for identifying translation start sites in *E. coli* bacteria genome with positive and unlabeled samples.
- Created a semi supervised neural network using the [Non-negative risk estimator](#) framework under the PU learning paradigm to address dataset constraints. Able to identify **60%** of masked start sites in the test set with a recall of **12%**.

**Research Assistant** **Bioinformatics Research Center, [Nanyang Technological University](#)**  
*Singapore* **Jul 2018 – Dec 2018**

- Advised by Dr. Kwoh Chee Keong. Area of research: Machine Learning, Drug Discovery. Outcomes: [Thesis](#) and [Poster](#)
  - Leveraged Variational Graph Autoencoder for drug-target interaction prediction, achieving **89%** AUC (near state-of-the-art).
  - Predicted Virulence (LD<sub>50</sub>) level of Influenza strains using traditional (GBDT, SVM, etc.) & GCN-based methods with **68%** acc.
  - Worked on improving generative models such as [DruGAN](#) for de novo drug generation using Tensorflow.
- 

## INDUSTRY EXPERIENCE

**Data Science Intern** **Faire**  
*San Francisco, CA, USA* **May 2022 – Aug 2022**

- Deployed an ML model (XGBoost) and heuristics for predicting users with no payment intent to reduce underwriting risk and proposed actions to take on users based on the model score.

- The model directly addressed **40%** of Faire’s defaults and showed a **25%** increase in commission over the baseline.
- Developed an end-to-end pipeline using SQL for preprocessing & feature selection and Python for model training & analysis.
- Designed optimization framework to select the best heuristics to aggregate model results for users across orders.
- Created a heuristic queue to identify fraud brands based on reviews, shipping info, and connected users with **65%** F1 Score.

## Data Scientist, Software Engineer

Bengaluru, KA, India

**Razorpay**

Jun 2019 – Aug 2021

### Payments Data Team

- Devised a novel recommender system using users’ historical transactions to personalize the Razorpay checkout page by ranking payment methods. Observed an increase in user conversion rate by **4%** in backtest.
- Developed a model testing pipeline using Apache Spark to test the model on extensive data with over 100 million samples.
- Implemented a real-time prediction REST API around the model using Scala and Akka HTTP with a latency less than 300ms.

### E-Commerce Fraud Detection Team

- Pioneered the development of critical components within Thirdwatch, a fraud detection platform, encompassing core business logic, ML models, the Data Science platform, and supporting backend components and infrastructure (used AWS).
- Created and deployed an ensemble XGBoost + CNN model to predict if an address is incomplete for delivery with **94%** AUC.
- Trained XGBoost models for predicting fraud e-commerce orders with **75%** AUC and deployed using H2O AutoML & AWS EC2.
- Implemented & scaled pipelines for model training, deployment, monitoring, and data ingestion using MLOps tools & Airflow.
- Published a workshop paper on shipping address classification at the E-Commerce workshop in **SIGIR 2021**.

## TEACHING EXPERIENCE

### Part-Time Lecturer

Courses: Data 101

**Rutgers University**

Fall 2021

### Teaching Assistant

Courses: Neural Networks, Digital Design, Computer Networks

**Birla Institute of Technology and Science, Pilani**

Fall 2017, Spring 2019

## SKILLS

**Expertise:** Machine Learning, Deep Learning, Natural Language Processing, Graph Representation Learning, Computer Vision, Drug Discovery, Statistical Models, Hypothesis Testing, Software Engineering, Multi-GPU Inference and Training

**Languages:** Python, Scala, SQL, C++, Java, Go, Bash (Linux), R, LaTeX

**ML & Visualization:** PyTorch (+ Lightning, Geometric, Torchvision, Torchtext), Tensorflow, Keras, Scikit-learn, MLLib, OpenCV, NLTK, Pandas, NumPy, SHAP, MLxtend, Matplotlib, Seaborn, Plotly

**Big Data & Backend:** PySpark, Ray, MySQL, Snowflake SQL, MLFlow, Flask, Docker, Airflow, HDFS, MapReduce, AWS (S3, EC2, Sagemaker), Elasticsearch

**Computational Biology & Chemistry:** RDKit, DeepChem, Molecular Operating Environment (MOE), MDAAnalysis, Auto-Dock Vina

## NOTABLE PROJECTS

**Visual Question Answering with Vision-Language Pretraining** [\[Code\]](#) [\[Paper\]](#) **Rutgers | 2023**

- Achieved exceptional accuracy (direct: **50%**, multiple-choice: **57.7%**) on the challenging [AOKVOA](#) dataset using the [GIT](#) model.

**Unsupervised Question Answering, Fact-Checking, and Entity Linking using Retrieval** [\[Code\]](#) [\[Paper\]](#) **Rutgers | 2022**

- Utilized the T0 language model for unsupervised text retrieval by re-ranking retrieved passages based on their prob scores.

**Abstract Visual Reasoning using Vision Transformers** [\[Code\]](#) [\[Paper\]](#) **Rutgers | 2022**

- Assessed the effectiveness of ViT and BEiT pre-trained transformer models in solving the [RPM IQ test](#) utilizing [L-RAVEN](#) dataset.
- Attained a significant **71.5%** test accuracy by improving and fine-tuning a ViT variant, highlighting its competitive performance.

## PUBLICATIONS

- **Waradpande, V.A.;** Prakash, P.V.S.; Jhaveri, N; Agarwal S ”Predicting Completeness of Unstructured Shipping Addresses Using Ensemble Models” [**SIGIR eCom ’21**] ([PDF](#))
- **Waradpande V.A.**, “Applications of Deep Graph Matrix Completion Models in Bioinformatics” [**Undergraduate Thesis**] ([PDF](#))