# PERSONAL PROFILE

Data Scientist with a strong programming background and experience in big data, machine learning, deep learning, and traditional statistical techniques. Passionate about finding meaningful ways data can help people.

# RECENT WORK EXPERIENCE

**Bioinformatics Researcher with Dr. Genevera Allen** **August 2016 – Current**

* Collaborating on building a predictive model of Alzheimer’s disease to beat clinical predictions using deep learning and traditional machine learning on genomics data. Data consists of SNPs, MicroRNA data.

**Freelance Data Scientist and Shiny Developer** **June 2015 – Current**

* Created a Shiny module for CloudDataAnalysis.com to provide an easy way for medical professionals to run longitudinal analyses in the cloud.
* Ran a statistical analysis of a knee-surgery study with Dr. Lionberger at the Texas Medical Center to determine statistical significance

**Software Engineer and Algorithms Intern with Buoy Inc.** **June 2016 – August 2016**

* Improved user-perceived accuracy of results from 50% to 70% and internal medical accuracy from 80% to 90% with a belief change simulation algorithm.
* Worked on a team with the Django and React frameworks to create a new feedback feature for interactive medical diagnostic product for both consumers and medical professionals.

**Data Science Intern with UCD Davis iAMSTEM Group** **June 2015 – August 2015**

* Created a logistic regression model to predict student probation and dismissal using admission datafrom 15 years of student data and wrote up policy recommendation in line with the results.
* Developed multiple data dashboards to visualize over 4.4 million student records.
* Computed and visualized a novel measure of course difficulty that built upon established literature.

**Intern at Imagine Easy Solutions LLC (EasyBib.com)**  **June 2014 – August 2014**

* Created a mathematical model of the advertising demand stack to predict better pricing points

# TECHNICAL SKILLS

* **Kaggle Competition Expert** <https://www.kaggle.com/andrewdumit>
* **Github Profile:** <https://github.com/adumit>
* **Software and Programming:** Python (PySpark, Django, scikit-learn, numpy, scipy, pandas), R, SQL, Spark, Hadoop MapReduce, Microsoft Excel, Git Version Control, React, AWS
* **Machine Learning:** Deep learning, classification, regression, clustering, dimension reduction
* **Statistical Methods:** Time series, regression models, hypothesis testing and confidence intervals
* **Selected Coursework Topics:** Deep Learning, Decision Forests, Gradient Boosting, Stochastic Processes, Recommender Systems, Autoencoder dimension reduction

# EDUCATION

**Bachelor of Arts in Statistics, Rice University, Houston, Texas Expected Graduation 2017**

* GPA 3.95

**AWARDS**

* 3rd Place in University College London Data Science Student Challenge 2016
* President’s Honor Roll, Spring 2014