EMAD SIDDIQ

emadsiddiq@berkeley.edu | [Website](#) | +1 (510) 990-5437 | Berkeley, CA

EDUCATION

University of California, Berkeley   **May 2021**
Bachelor of Arts in Data Science and Political Economy   Berkeley, CA
                                                                                          GPA: 3.24

Relevant Courses: Machine Learning, Design and Analysis of Algorithms, Stochastic Processes, Principles and Techniques of Data Science, Probability for Data Science, Structure and Interpretation of Computer Programs, Data Structures and Algorithms, Linear Algebra, Econometric Analysis, Ethics of Data Science, Political Economy of Silicon Valley, Information, Media and the Public Discourse

Frameworks: React, Flask, Scikit-Learn, TensorFlow, NLTK, PyTorch, Fetch API, networkX, NumPy , pandas, Ruby on Rails, MaterialUI, AWS, Apache Spark

Languages: C, Java, R, Python, SQL, Julia, C++, C#, Scheme, HTML, CSS, Javascript, PHP, Ruby, Objective-C

EXPERIENCE

Undergraduate Research   Berkeley, CA
UC Berkeley Law School   May 2020 – August 2020

- Conducted research for Professor Sonya Katyal's Trademark & Branding Project as the team's lead data scientist
- Parallelized data extraction from remote XML files, with limited cloud storage (5 GB), for over 100GB of data
- Designed a fast matching algorithm for relevant companies from a list provided by Haas School of Business

PROJECTS

Pashto dictionary:
- Undertook a semester long research project to understand digital linguistic inequality, and concluded that NLP based techniques can aid digitization of at risk languages like my mother tongue Pashto
- Implemented a Trie data structure & a simultaneous search algorithm for both Pashto & Latin scripts
- Designed an asynchronous website, offering a rich yet accessible experience for data capped Pashtun users

Good News
- Tackled the flood of negative news following the pandemic, by gathering good news from 50+ news sources
- Applied Sentiment Analysis using Python & NLTK, to rank the positive of news articles from NewsAPI
- Developed a mobile first and human friendly dynamic design, with an emphasis on an intuitive UI\UX

Tech-Khwani
- Leveraged Google Colab's free resources to teach CS fundamentals with no prerequisites to Pakistani students
- Created easily accessible notebooks with a contextualized pedagogical approach, for the Data Science lifecycle
- Tutorials included: training a neural network to predict housing prices in Islamabad, & predicting passenger survival on the Titanic voyage

Data Mining
- Mined 10 years archives for Pakistan's leading newspaper Dawn.com, using BeautifulSoup on a remote server
- Mined over 10GB of OECD's PISA 2018 data; a massive survey about the state of education around the world