**Mohammad Shahebaz** Mobile: +91 9177397148

Top 1% Kaggle Master, Scikit-learn contributor Email: shahebaz13@gmail.com

[github.com/shaz13](http://www.github.com/shaz13)

[kaggle.com/shaz13](http://www.kaggle.com/shaz13)

**EDUCATION**

**---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

* **Jawaharlal Nehru Technological University**                               Telangana, India

Bachelor of Technology in Electronics and Communications, 72.3%                      Aug. 2015 – May 2019

**ACHIEVEMENTS**

**---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

* **International Achievements**
* **Top 1%** out of **1.02 million** usersin competitions worldwide on Kaggle, 2019
* **Top 1%** out of **7198** teams in Kaggle’s Home Credit Default Risk Challenge, 2018
* **Top 3%** out of **4047** teams in Quora Insincere Questions Classification at Kaggle, 2019
* India finalist from the south region for **Microsoft Office World Championship**, Delhi, 2013
* Regional Gold medalist of **International Math Olympiad**, 2008
* **National Achievements**
* **Winner** of most innovative solution at Brainwaves 2019 by Societie Generale, Bangalore, 2019
* **Winner** of Analytics Vidhya Hiring Hackathon out of 1410 applicants, 2018
* **1st Runner-up** out of 9107 teams at Code Gladiators grand finale in Machine Learning theme, 2018
* **Top 10** finalist for Master Orator Championship by Toastmaster International, Hyderabad, 2016

**EXPERIENCE**

**---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

* **H2O.ai** Chennai, India

Data Scientist Intern                                                                                                       Jan 2019 – Present

* **Driverless A.I**:  Validate engineered features on real world datasets and live International competitions
* **Product Development:**Implementing BERT Classifier in TensorFlow for L.P in Driverless A.I
* **Research:** Improvising benchmarks of state of art NLP classification and sentiment tasks
* **Analytics Vidhya**    Gurgaon, Haryana

Data Science Intern                                                                                                        May 2018 – Nov. 2018

* **Data Hack Platform**: Fra­­med time series challenge for Sapient Talent Hunt Competition
* **Product Development:**Authored inhouse machine learning library for automated feature engineering
* **Evangelism:**Guest speaker for Hyper Connected World sponsored by Intel at IIIT Hyderabad
* **Oppia.org,** Google Open Source Project  GitHub, Remote

Social Committee Lead, Member Mar. 2017– May. 2018

* **Development:**Resolved critical design issues and improved the landing page performance by **46.1%**
* **MapReduce:** Resolved a major security bug in backend with MapReduce job iterating over **400000 users**

**PROJECTS**

**---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

* **Telugu Text Summarization:** Assisting a research paper on summarization using semi-supervised methods on Telugu language. To be published at Scopus Index Journals
* **Toxic Classification:** Top 1% **AUC 0.9852** score. Applied various embeddings and used various ensemble models with stacked layer of Logistic Regression
* **Deprecation of English stop words:** Contributed to scikit-learn on cross-validated prediction documentation and raised a cycle for deprecation of English stop words
* **Image Captions Generation:** Image captions generation using ResNet-152 encoder and LSTM decoder