- This internship assignment is from September until the end of December 2017. Will you be available to be on-site 40 hours a week?

  Yes, I would be available during that period for the internship

---

- Why are you interested in this position?

  This position perfectly aligns with my academic background, skill set and my professional goals. The internship involves designing intelligent systems to make data driven decisions and presents responsibilities that I have worked on in the past and have experience in. It also provides me with the opportunities to put my skills to test while learning new ones along the way.

---

- What level of experience do you have with data analysis?

  I have 2 years of hands-on programming experience in R and python. In this duration, I had the opportunity to work with some of the most popular machine learning and statistical packages. I have interned at HoldingWilley, one of India's most popular cricketing website as a Data Analyst, I have also worked as a Research Assistant in Machine Learning at the University of Mumbai. Additionally I have worked on several interesting data science and computer vision projects mentioned below.

---

- How would you define data science?

  Personally, I believe data science is extracting userful information and insights from a large collection of unstructured and uninformative data to make informed decisions.

---

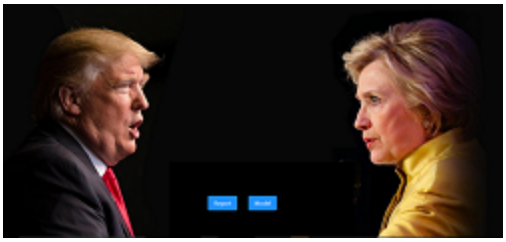
- Describe a project you've completed that required data analysis

  Recently, I worked on the following projects.

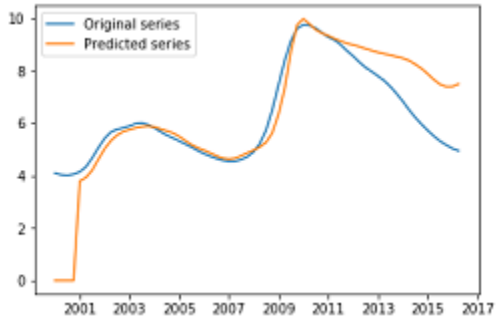
  - Predicting amazon product backorders .
    - The goal of this project was to determine whether a product would go on backorder or not.
    - There were two major issues with the dataset
      - Target classes were imbalanced [98% samples belonged to 'no backorder']
      - Target classes were non-separable
    - I achieved 90% accuracy and 85% recall using the following techniques
      - Non linear dimension reduction and visualization, support vector machines and gradient boosting ensemble learner
    - The model was designed in R.

  ---

  - United States Presidential Elections- predicting voter preference
    - The goal of this project was to determine individual's voting preference given his demographic information and his policy alignment.
    - I designed two ensemble learners, a gradient boosting model and a random forest model. The gradient boosting model achieved 90% accuracy on the test set.
    - The model was designed in R and the report was designed in Tableau
    - Link to Project

- 

---

- Predicting unemployment rate using long short term memory recurrent neural networks
  - The goal of this project was to predict unemployment rate in the United States.
  - The Gross Domestic Product time series followed an inverse time lagged relationship with the unemployment time series, thus the GDP series was used as the source series.
  - The model's biggest challenge was to adapt to the inconsistent oscillations in the target series.
  - My model achieved an RMSE error of less than 0.05 and did an excellent job in following the target series
  - The model was developed in python using keras deep learning repository with tensorflow backend
  - Link to Project
  - 


---

- In what ways have you worked collaboratively with a team before?

During my tenure as a Research Assistant, I worked with 2 fellow researchers. Our work ethic and collaboration lead to the project being completed well before its due date. Additionally, I served as the General Secretary of Computer Society of India, student chapter. Here I had the opportunity to lead a team of 16 people. My responsibilities included managing the daily activities of the chapter, organizing workshops, setting up assignments and deadlines for the team and ensuring that these were met. I am proud to have led the team to win the Best Student Chapter award in my tenure. I am currently leading the Association of Information Science and Technology chapter at Indiana University.

---

- What programming experience do you have?

I hold a bachelor's degree in Computer Science and have 2 years of programming experience in R and python. I have worked with the following machine learning and statistical packages

- Machine Learning
  - Caret
  - MlR
  - Scikit-learn
- Visualization
  - GGplot
  - Matplotlib
  - Seaborn

- Neural networks/ Deep learning
  - Keras
  - Lasagne
  - Mxnet
- Frameworks
  - h2o
  - tensorflow

Additionally I have worked with Tableau, SAS and STATA for my academic projects. I have utilized Indiana University's high performance cluster computing resources to deploy my code in a parallel environment.

---

- What new skills have you recently acquired?

  Recently I have been working with shiny applications for interactive visualizations. I am working towards building a shiny application for the Presidential Elections project mentioned above. The goal is to replicate the tableau report, providing users with control over the visualizations in R.

---

- What skills are you interested in acquiring?

  I have worked with Python in a parallel computing environment, deploying my computer vision/deep learning projects on multiple CPU's and GPU's. I aim to learn to deploy R codes in a parallel environment. Additionally, having worked with standalone programs and projects, I am eager to learn to deploy these on scale using frameworks such as spark.

---

- What is your current work authorization status?

  I am currently residing in the US on a student visa [F1-B]. As part of the OPT Program, I am allowed to work for 3 years.