AI Skunkworks Research

Model Interpretability

Research Head:

Abhishek Maheshwarappa

Email: maheshwarappa@northeastern.edu

LinkedIn

Expected Start Date: 08/07/2020

How to get started:

1.   Signup using the form shared with all the members.

2.   We will reach out with a challenge.

3.   Complete the challenge within a given time

4.   Come to the weekly meetings that starts after the first week of September (7th September)

Prerequisites:

1.   Experience with Python or R

2.   Worked and implemented ML Models in projects

3.   Complete the challenge given by us

Abstract:

As technology is expanding into various domains right from shopping to academics to and other niche domains such as house cleaning - it is significantly impacting our lives. Today, there is a surge in data all around, and the use of machine learning as gone up exponentially which is hugely impacting life. Making a machine trustworthy and reliable is one of the most important goals of data science today. Models are many times used as black boxes, wherein we give input, know little of what happens inside the model, and get an output. To understand current work and implement them on a wide variety of problems. Identify the problems in the current research with evidence and try to improve it by introducing a new technique for interpretability and write a paper on it. Introducing a new method for explaining to models so that humans can interpret the model in an easier way. The field has a lot of scopes to explore and research around, google is spending lots of money to work on interpretability, companies are looking explainable AI, so this is the right time to invest to undertake research in this field.