



Data Systems Group

XAI Spring2020 Workplan – Abdul Wahab



Scope: A Comprehensive Qualitative and Quantitative Evaluation of Model-Agnostic Local Interpretability Methods.

Phase 1: Benchmarking of Tabular Interpretability methods on 3 BioMedical Tabular Datasets		
Task	Comments	Timeline
1. Consolidating Metrics + Implementing New Metrics	Consolidating all existing and relevant metrics for interpretability + implementing any new metrics that suit the Tabular Dataset type Interpretability.	9-15 th March
2. Dataset and Model setup	Gathering the relevant datasets and setting up the model and interpretability methods on these datasets.	16-22 nd March
2. CLEAR Benchmarking	Understanding the approach behind this approach method and doing the benchmarking experiments	23-29 th March
4. BreakDown Benchmarking	// //	30 th March-05 th April
5. D-LIME Benchmarking	// //	06 th April-12 th April
6. Counterfactual Explanations Guided by Prototypes	// //	13 th April - 19 th April
Phase 2: Continued Benchmarking of 3 more Methods		
7. LORE	// //	20 th April-26 th April
8. Contrastive-Explanation-Method CEM	// //	27 th April - 03 rd May
9. Contrastive Explanation (Foil Trees)	// //	04 th May - 10 th May