Data Systems Group

Institute of Computer Science, University of Tartu

XAI-Spring2020 Project Plan
5th March 2020
Abdul Wahab

Scope: A **C**omprehensive **Q**ualitative and **Q**uantitative **E**valuation of **M**odel-**A**gnostic **L**ocal **I**nterpretability **M**ethods.

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 andy new metrics that suit the Tabular Dataset type Interpretability.	9-15th March
2. Dataset and Model setup	Gathering the relevant datasets and setting up the model and interpretability methods on these datasets.	16-22nd March
2. CLEAR Benchmarking	Understanding the approach behind this approach method and doing the benchmarking experiments	23-29th March
4. BreakDown Benchmarking	// //	30th March-05th April
5. D-LIME Benchmarking	// //	06th April-12th April
6. Counterfactual Explanations Guided by Prototypes	// //	13th April - 19th April
Phase 2: Benchmarking of Text Interpretbility Methods on 3 Datasets		
7. Consolidating Metrics + Implementing New Metrics	Consolidating all existing and relevant metrics for interpretability + implementing andy new metrics that suit the Tabular Dataset type Interpretability.	20th April-26th April
8. Dataset and Model setup	Gathering the relevant datasets and setting up the model and interpretability methods on these datasets.	27th April - 03rd May

9. POLAR	Understanding the approach behind this approach method and doing the benchmarking experiments	04th May - 10th May
8. Text Dataset Interpretability Method 2	// //	11th May – 17th May
9. Text Dataset Interpretability Method 3	// //	18th May – 24th May
Phase 3: Benchmarking of Above Methods on Images		
7. Consolidating Metrics + Implementing New Metrics	Consolidating all existing and relevant metrics for interpretability + implementing andy new metrics that suit the Tabular Dataset type Interpretability.	15th May – 24th May
8. Dataset and Model setup	Gathering the relevant datasets and setting up the model and interpretability methods on these datasets.	25th May - 31st May
10. Counterfactual Explanations Guided by Prototypes	Understanding the approach behind this approach method and doing the benchmarking experiments	1st June - 6th June
11. Contrastive Explanation (Foil Trees)	// //	7th June - 13th June
12. Convex Density Constraints for Computing Plausible Counterfactual Explanations	// //	14th June - 20th June