A framework for benchmarking discrimination-aware models in machine learning

**ABSTRACT**

Discrimination-aware models in machine learning are a recent topic of study that aim at minimizing the adverse impact of machine learning decisions for certain groups of people due to ethical and legal implications. We propose a benchmark framework for assessing discrimination-aware models. Our framework consists of systematically generated biased datasets that are similar to real world data, created by a Bayesian network approach. Experimental results show that we can assess the quality of techniques through known metrics of discrimination, and our framework is flexible and can be extended to most real datasets and fairness measures to support a diversity of assessments.

**EXISTING SYSTEM**

Discrimination-aware models in machine learning are a recent topic of study that aim at minimizing the adverse impact of machine learning decisions for certain groups of people due to ethical and legal implications. Discrimination-aware learning is a topic of research that aims at minimizing the impact of bias against certain groups due to ethical reasons and legal implications. Given a labeled dataset whose records represent individuals, let its attributes be divided into non-protected and protected attributes, such as race and gender. The problem consists of building a classifier that takes the non-protected attributes of an individual and maps them to a class label, so that the classifier has minimum discrimination and maximum accuracy.

**Disadvantages of Existing System:**

1. The problem consists of building a classifier that takes the non-protected attributes of an individual and maps them to a class label, so that the classifier has minimum discrimination and maximum accuracy.

**PROPOSED SYSTEM**

We propose a benchmark framework for assessing discrimination-aware models. Our framework consists of systematically generated biased datasets that are similar to real world data, created by a Bayesian network approach. The main contribution of this work is a framework for comparing discrimination-aware learning models. To the best of our knowledge, there is no other such benchmark framework in the literature. Our framework comprises systematically generated biased datasets that are sampled from Bayesian networks learned from real world data. Our main concern is to explore alternative discrimination scenarios, that is, we want to learn from data representing different levels of bias for the purpose of analyzing the behavior of techniques when bias levels differ.

**Advantages of Proposed system:**

1. Our framework is flexible and can be extended to most real datasets and fairness measures to support a diversity of assessments.

# Hardware Requirements:

# Processor - Pentium –IV

* Speed - 1.1 GHz
* Ram - 256 MB
* Hard Disk - 20 GB
* Key Board - Standard Windows Keyboard
* Mouse - Two or Three Button Mouse
* Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - Java