

1. Pendahuluan

1.1. Latar Belakang

Bank Rakyat Indonesia (BRI) membangun loyalitas nasabah salah satunya dengan melakukan akuisisi atau pengambilan kepemilikan perusahaan berupa aset untuk menjadi agen (pihak yang diajak bekerja sama). Agen BRILink yang tersebar di seluruh Indonesia merupakan penerapan akuisisi agen di BRI. Dengan jumlah nasabah BRI yang besar, hal ini dapat menimbulkan peluang terjadinya akuisisi agen BRILink yang tidak adil. Ketidakadilan ini dapat menyebabkan bias sehingga perlu dilakukan pembelajaran dalam suatu data. Bias dapat muncul disebabkan oleh fitur-fitur, sebagaimana [1] menunjukkan bahwa jenis kelamin dapat menyebabkan bias pada data. Bias yang muncul pada sebuah dataset, akan menyebabkan individu atau grup dirugikan [2]. Terdapat beberapa fitur demografi dalam data BRILink yang dapat menimbulkan bias, seperti jenis kelamin, usia, lokasi, pekerjaan, dan ras [3-5]. Disebut fitur demografi, karena fitur tersebut memiliki banyak nilai yang beragam sehingga dapat mengindikasikan bias. Misalnya pada penelitian [6] ras digolongkan sebagai fitur demografi karena, ras dapat menentukan seberapa berat hukuman yang diberikan berdasarkan histori kriminal yang ada. Jika histori kriminal condong pada orang berkulit hitam maka orang berkulit hitam akan mendapatkan hukuman yang lebih berat, hal ini menunjukkan suatu diskriminasi. Merujuk pada dataset BRILink, agen yang diakuisisi lebih sedikit dibandingkan dengan agen yang tidak diakuisisi, hal ini menunjukkan bahwa *fairness* belum tercapai.

Oscar, Deho et.al (2021) mengimplementasikan algoritma mitigasi pada dataset *learning analytics*, dan menunjukkan bahwa dataset menghasilkan prediksi yang lebih baik ketika data *fair*. *Fair* adalah keadaan dimana tidak adanya diskriminasi pada data. Oleh karena itu mitigasi bias perlu dilakukan pada dataset BRILink. Untuk melakukan mitigasi bias terdapat suatu *toolkit* bernama AI Fairness 360 (AIF 360) yang merupakan *toolkit* terbaru dan lengkap karena AIF 360 juga menyediakan algoritma untuk deteksi bias [7]. Deteksi bias pada dataset BRILink dilakukan dengan metode *disparate impact* (DI) dan *statistical parity difference* (SPD) karena metode ini sederhana untuk dikomputasikan. Setelah dilakukan pendeteksian bias, dilakukan mitigasi bias melalui *pre-processing dengan metode reweighing* dan *learning fair representation*, *in-processing dengan metode prejudice remover* dan *adversarial debiasing*, dan *post-processing dengan metode calibrated equalized odds* dan *reject option classification* [8].

Hasil deteksi bias dengan DI dan SPD sebelum dan sesudah melalui proses mitigasi akan dibandingkan untuk melihat metode yang paling efektif dalam melakukan mitigasi pada dataset BRILink. Penelitian ini dapat mengatasi bias pada data dan menghasilkan *fairness* prediksi untuk akuisisi agen BRILink. Penelitian ini tentunya akan berbeda dengan penelitian-penelitian terdahulu [7], [9]-[12] terkait *bias and fairness*, karena pada penelitian ini menggunakan dataset BRILink dan kombinasi metode AIF 360 yang berbeda dengan penelitian sebelumnya.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah penelitian ini adalah bagaimana mendeteksi bias yang dilakukan dengan metode DI dan SPD serta bagaimana melakukan mitigasi bias dengan metode *reweighing*, *learning fair representation*, *prejudice remover*, *adversarial debiasing*, *reject option classification*, dan *calibrated equalized odds* sehingga dapat menghasilkan data yang lebih *fair*.

1.3. Tujuan

Adapun tujuan dari penelitian untuk tugas akhir ini adalah melakukan deteksi bias dengan DI dan SPD serta menerapkan metode mitigasi bias yang ada pada tahap *pre-processing*, *in-processing*, dan *post-processing* dari AIF 360 untuk menghasilkan prediksi yang lebih *fair* dalam akuisisi agen. Hasil deteksi bias sebelum dan sesudah mitigasi akan dibandingkan untuk menemukan metode mitigasi bias yang paling efektif untuk mengatasi bias.