

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam kehidupan modern, data menjadi pusat dalam kehidupan manusia. Perkembangan sains, teknologi dan budaya yang berkembang sangat pesat didasarkan pada data. Kegiatan bisnis dan ekonomi juga semakin mengendalikan ketersediaan data-data yang ada. Bahkan, dinamika budaya, seni dan sosial-politik tidak terlepas dari data. Diplomasi dan negosiasi internasional makin bergantung pada data. Sebagian aspek religiusitas dan spiritualitas pun nampaknya memiliki porsi yang cukup besar dan kuat atas data.

Indonesia sangat memerlukan *Data Science* untuk memecahkan berbagai tantangan besar dan untuk membuat lompatan ke depan, untuk menunjukkan kemampuan dan kesungguhannya dalam membangun reputasi nasional maupun internasional. *Data Scientist* sudah sangat diperlukan Indonesia saat ini dan terus diperlukan dalam jumlah yang semakin besar. Ini menjadi tantangan bagi kalangan pemerintah, penyelenggara layanan publik, berbagai organisasi, termasuk sekolah-sekolah dan universitas. Di masa depan, *Data Scientist* dapat menjadi salah satu pekerjaan yang sangat dibutuhkan pada suatu perusahaan [1].

Salah satu tantangan yang dihadapi perusahaan di masa sekarang adalah pelanggan yang tidak kembali berbelanja di *platform* perusahaan. Hal ini dapat diakibatkan oleh berbagai faktor dan dalam pelaksanaan analisisnya memerlukan waktu dan biaya yang tidak sedikit. *Data Science* dapat digunakan untuk menganalisis sebab pelanggan tidak kembali berbelanja pada suatu perusahaan dengan cepat dan mudah sehingga dapat meminimalisir kerugian dan memaksimalkan profit perusahaan. Hal inilah yang mendorong penulis untuk menganalisis permasalahan ini.

Penelitian ini berfokus pada *churn customer analysis* yaitu metode yang digunakan oleh perusahaan untuk mengidentifikasi dan memprediksi pelanggan yang berpotensi untuk tidak kembali lagi bertransaksi di perusahaan. Penulis

menggunakan data yang berasal dari situs Kaggle dengan nama *Fashion Campus* [2]. Peralatan yang digunakan adalah menggunakan Google Colab dan bahasa pemrograman *Python*. Penelitian ini menggunakan tiga model algoritma yaitu *Logistic Regression*, *Decision Tree* dan *Random Forest Classifier*. Beberapa algoritma perlu dipakai untuk mengurangi bias dan mencari algoritma terbaik dalam memprediksi *churn* pelanggan. *Assessment* yang dipakai dalam mencari algoritma terbaik adalah menggunakan *recall*.

1.2 Rumusan Masalah

1. Apa *feature-feature* terbaik yang dapat digunakan untuk memprediksi *churn* pelanggan pada penelitian ini?
2. Apa algoritma terbaik yang dapat digunakan untuk memprediksi *churn* pelanggan pada penelitian ini?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini mencakup beberapa hal di antaranya:

1. Menggunakan tiga metode algoritma yaitu *Logistic Regression*, *Decision Tree* dan *Random Forest Classifier*.
2. Software yang digunakan adalah Google Colab dan bahasa pemrograman *Python*.
3. Data yang digunakan berasal dari situs Kaggle dengan *dataset* fiktif.

1.4 Tujuan Penelitian

1. Untuk mengetahui *feature-feature* terbaik yang dapat digunakan untuk memprediksi *churn* pelanggan pada penelitian ini.
2. Untuk mengetahui algoritma terbaik yang dapat digunakan untuk memprediksi *churn* pelanggan pada penelitian ini.

1.5 Manfaat Penelitian

Dalam melakukan penelitian ini terdapat beberapa manfaat yang diketahui oleh penulis seperti bagaimana memprediksi *churn* pelanggan dengan efektif menggunakan *software-software* tertentu, *feature-feature* terbaik apa saja yang dapat digunakan dan algoritma terbaik yang dapat digunakan. Nantinya, penulis berharap di kemudian hari, peneliti lain dapat mencari *feature*, metode dan algoritma terbaik dalam memprediksi *churn* pelanggan berdasarkan referensi dari penelitian ini.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN, berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka dan dasar-dasar teori yang digunakan pada penelitian.

BAB III METODE PENELITIAN, di dalamnya terdapat alur penelitian serta alat dan bahan yang digunakan pada penelitian.

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan tahapan yang penulis lakukan dalam mengembangkan *dataset*, *training* dan *test* hingga melakukan *assessment* model di objek penelitian.

BAB V PENUTUP, berisi kesimpulan dan saran yang dapat penulis rangkum selama proses penelitian.