

# **Pendekatan Data Science Untuk Menemukan Churn Pelanggan Pada Sector Perbankan Dengan *Machine Learning***

## **BAB 1 PENDAHULUAN**

### **1.1. Latar Belakang**

Peralihan pelanggan merupakan fenomena dimana pelanggan perusahaan berhenti membeli atau berinteraksi dengan perusahaan. Churn rate yang tinggi berarti banyak pelanggan yang tidak mau membeli barang atau jasa dari perusahaan [1]. Churn pelanggan adalah perhitungan matematis dari persentase pelanggan yang tidak mungkin melakukan pembelian lagi dari perusahaan. Menghindari gesekan pelanggan adalah salah satu faktor kunci dalam memaksimalkan profitabilitas organisasi seperti bank [2]; [3]; [4]; [5]; [6] dan layanan telekomunikasi [7]; [1]; [8]. Di sektor perbankan, inovasi teknologi saat ini memberikan kemudahan bagi pelanggan untuk membuka rekening bank baru dan mentransfer semua aset bahkan tanpa harus datang dimana pelanggan dapat melakukannya dari rumah. Situasi ini membuat bank lebih tertarik pada topik loyalitas pelanggan. Namun, sebelum memulai cara yang efisien untuk mempertahankan pelanggan yang sudah ada, sangat perlu memprediksi pelanggan mana yang akan berhenti[9].

Mengakhiri hubungan pelanggan dengan perusahaan memiliki nilai yang tidak dapat disangkal untuk semua organisasi karena memungkinkan mereka untuk menyiapkan kampanye bertarget untuk mempromosikan loyalitas pelanggan [9]. Prakiraan churn yang akurat secara efektif mendukung strategi loyalitas pelanggan dan merencanakan kampanye pemasaran ekonomi, menghasilkan penghematan yang signifikan bagi penyedia layanan. Untuk bertahan dari persaingan yang ketat ini, perusahaan telekomunikasi menjadi lebih agresif dengan berinvestasi lebih banyak dalam mengembangkan data mining dan model berbasis pembelajaran mesin untuk analitik, peramalan, dan manajemen churn [7]. Banyak penelitian telah menemukan bahwa teknologi pembelajaran mesin sangat efisien dalam memprediksi situasi ini. Berbagai pendekatan pembelajaran mesin untuk prediksi churn telah diusulkan dalam literatur.

Namun, tugas ini sangat sulit karena distribusi kelas yang tidak merata dan jumlah kelas pelanggan yang tidak dibatalkan melebihi jumlah kelas churn[8]. Belajar dari distribusi kelas yang tidak seimbang sangat kompleks dengan sebagian besar algoritma pembelajaran mesin

klasik, karena cenderung mengklasifikasikan sebagian besar kelas dengan benar dan mengabaikan yang langka. Masalah besar lainnya adalah ada banyak faktor yang mempengaruhi gesekan pelanggan dan hubungan di antara mereka sangat kompleks [10]. Mengukur peran faktor-faktor ini juga merupakan tugas yang kompleks[10]. Oleh karena itu, interpretasi model pembelajaran mesin sangat penting. Sebagian besar studi literatur sebelumnya berfokus pada peningkatan akurasi model dan kurang tertarik untuk memahami model yang dihasilkan dan mengukur peran faktor-faktor yang mempengaruhi keakuratan klasifikasi. Makalah ini membahas tentang penerapan data science dalam menyajikan analisis dan menemukan churn pelanggan pada sektor perbankan.

Banyak pendekatan yang diterapkan untuk memprediksi churn di perusahaan. Sebagian besar pendekatan ini telah menggunakan pembelajaran mesin dan penambangan data. Sebagian besar pekerjaan terkait berfokus pada penerapan hanya satu metode penambangan data untuk mengekstrak pengetahuan, dan yang lainnya berfokus pada membandingkan beberapa strategi untuk memprediksi churn.

## **1.2. Rumusan Masalah**

Rumusan masalah yang akan dibahas dalam penelitian ini adalah bagaimana cara menemukan churn pelanggan pada sector perbankan dengan machine learning dengan pendekatan data science?

## **1.3. Tujuan dan Manfaat**

Tujuan penyelesaian masalah pada penelitian ini adalah untuk menemukan churn pelanggan pada sector perbankan dengan Machine Learning dengan pendekatan data science.

Sedangkan manfaat pada penelitian ini adalah :

1. Menganalisis dan mencari keadaan churn pelanggan
2. Melakukan segmentasi pelanggan
3. Memprediksi jumlah persentase nasabah yang keluar dari bank dengan analisis regresi
4. Menemukan pelanggan yang memiliki potensi akan meninggalkan bank

## **1.4. Batasan Masalah**

Berikut beberapa batasan masalah dalam penelitian ini:

1. Kumpulan data churn banking bersumber dari kaggle <https://www.kaggle.com/santoshd3/bank-customers>
2. Menerapkan data science dalam menyajikan analisis dan menemukan churn pelanggan pada sektor perbankan.
3. Menggunakan pendekatan Machine Learning dan Data Mining.
4. Variabel yang digunakan adalah variable diskrit dan variable kategoris
5. Bahasa pemrograman yang digunakan adalah *Python*.

### **1.5. Keterbaruan**

1. Kepuasan pelanggan dan churn pelanggan telah diidentifikasi sebagai dua faktor yang berkontribusi terhadap keberhasilan industri dan oleh karena itu topik hangat diteliti di berbagai industri, seperti pariwisata [11], telekomunikasi [1] dan banking [12]; [4].
2. Ahmad et al [8] mengembangkan model perkiraan churn yang membantu operator memprediksi pelanggan mana yang kemungkinan besar akan churn. Model yang dikembangkan dalam makalah ini menggunakan teknologi pembelajaran mesin pada platform data besar dan dibangun di atas metode rekayasa dan pemilihan fitur baru. Pengukuran under-standard curve area (AUC) digunakan untuk mengukur kinerja model, dan diperoleh nilai AUC sebesar 93,3%. Kontribusi penting lainnya dalam makalah ini adalah penggunaan jaringan pelanggan sosial dalam model prediktif dengan mengekstraksi fungsi analisis jaringan sosial (SNA).
3. Ammar et al [1] mengusulkan strategi berbasis baru untuk memprediksi tingkat churn perusahaan telekomunikasi. model prediksi churn ensemble yang digabungkan dengan pemodelan uplift berbasis biaya. Model ini menguntungkan organisasi dengan memenuhi tujuan bisnis secara efektif dan juga dengan mengurangi biaya secara signifikan jika dibandingkan dengan model lainnya. Pendekatan ensemble yang diusulkan meningkatkan proses prediksi dan memberikan akurasi 94,3%.
4. Almuqren et al [13] mengusulkan pendekatan baru dengan menggunakan penambahan media sosial untuk memprediksi churn pelanggan di bidang telekomunikasi. Metode baru yang diusulkan digunakan untuk mengekstrak umpan balik kepuasan pelanggan secara real-time dan digunakan untuk memprediksi churn pelanggan dengan tingkat akurasi 95,8%.

5. Kaur, Ishpreet, Kaur, Jasleen [5] menerapkan model pembelajaran mesin yang berbeda seperti *Logistic regression* (LR), *decision tree* (DT), *K-nearest neighbor* (KNN), *random forest* (RF) dan lainnya diterapkan pada dataset bank untuk memprediksi probabilitas pelanggan yang akan melakukan churn. Mereka melaporkan Random Forest berkinerja lebih baik di antara semua model yang digunakan.
6. Faris, H [14] mengusulkan model hybrid cerdas berdasarkan Particle Swarm Optimization dan jaringan saraf Feedforward diusulkan untuk prediksi churn. PSO digunakan untuk menyesuaikan bobot fitur input dan mengoptimalkan struktur jaringan saraf secara bersamaan untuk meningkatkan daya prediksi dan [15] menerapkan memprediksi perilaku pelanggan, untuk meningkatkan pemanfaatan gym dan retensi pelanggan.