Literature Review

Ada sejumlah besar penelitian terkait dengan Customer Relationship Management (CRM) yang mengatasi keterbatasan kumpulan data mereka dan mengusulkan berbagai algoritme dan teknik untuk mengidentifikasi segmentasi pelanggan dari kumpulan data tertentu. Subbagian berikut membahas pekerjaan terkait dari tantangan paling penting yang dihadapi para peneliti dengan solusi yang diusulkan masing-masing. Selain itu, kami akan memfokuskan tinjauan literatur kami untuk membahas pekerjaan terkait yang menggunakan teknik klasifikasi yaitu Logistic Regression, Decision Tree, Random Forest Tree

2.1 Customer Segmentation in Customer Relationship Management (CRM)

Segmentasi dapat dilihat sebagai penyederhanaan dari kerumitan berantakan berurusan dengan banyak pelanggan individu, masing-masing dengan kebutuhan yang berbeda dan nilai potensial [4]. Metode segmentasi pelanggan tradisional umumnya didasarkan pada metode klasifikasi pengalaman atau metode statistik sederhana [5]. Metode statistik tradisional mengelompokkan pelanggan menurut karakter perilaku sederhana atau karakter atribut seperti kategori produk yang dibeli atau wilayah tempat tinggalnya [6]. Metode segmentasi ini tidak dapat melakukan analisis yang lebih kompleks bahwa pelanggan seperti apa yang memiliki nilai potensial tinggi dan pelanggan seperti apa yang memiliki kredit tinggi. Dengan penerapan EC dan CRM yang ekstensif, perusahaan telah mengumpulkan semakin banyak data pelanggan. Teknik tradisional seperti regresi berganda tidak dapat mengatasi tingkat kerumitan ini. Akibatnya, keandalan dan validitas fungsi statistik yang digunakan untuk menghasilkan segmentasi atau untuk membangun model prediktif menjadi faktor kemungkinan penyebab ketidakpuasan pengguna CRM [7].

Data mining dapat dianggap sebagai metodologi dan teknologi yang dikembangkan baru-baru ini, menjadi terkenal pada tahun l994. SAS Institute mendefinisikan data mining sebagai proses memilih, mengeksplorasi dan memodelkan sejumlah besar data untuk mengungkap pola data yang sebelumnya tidak diketahui [5]. Dengan demikian, data mining dapat dianggap sebagai proses dan teknologi untuk mendeteksi hal-hal yang sebelumnya tidak diketahui untuk mendapatkan keunggulan kompetitif. Penambangan data menggunakan jaringan saraf, pohon keputusan, analisis tautan, dan analisis asosiasi untuk menemukan tren dan pola yang berguna dari data yang diekstraksi [6]. Penambangan data dapat menghasilkan wawasan penting termasuk model prediksi dan asosiasi yang dapat membantu perusahaan memahami pelanggan mereka dengan lebih baik. Banyak perusahaan besar saat ini memiliki terabyte data, di mana mereka mungkin dapat menemukan lebih banyak informasi tentang pelanggan, pasar, dan persaingan daripada yang mereka perlukan. Penambangan data memungkinkan pemasar untuk mengekstrak informasi bisnis yang berharga dengan lebih baik dari 'gunung data' dalam sistem perusahaan. Ini adalah solusi potensial untuk masalah besar yang dihadapi banyak perusahaan: data yang melimpah dan relatif kelangkaan staf, teknologi, dan waktu untuk mengubah angka dan catatan menjadi bermakna. informasi tentang pelanggan yang ada dan calon pelanggan. Data mining memungkinkan perusahaan untuk mengukur perilaku konsumen berdasarkan 100 atau lebih atribut, bukan tiga atau empat yang terkait dengan pemodelan statistik tradisional [7]. Semakin banyak atribut yang digunakan perusahaan, semakin besar kompleksitas data dan semakin besar kebutuhan akan alat penambangan data. Karena para praktisi dengan antusias mencari kelompok yang menguntungkan pelanggan yang loyalitasnya stabil, beberapa akademisi mulai mempertanyakan apakah segmen sebenarnya adalah entitas yang stabil dan banyak lagi fundamental apakah mereka benar-benar ada [7]. Metode segmentasi berdasarkan data mining yang dibuat oleh makalah ini dapat menyelesaikan permasalahan di atas karena model dapat mempelajari informasi baru yang dimasukkan kemudian dan mendapatkan aturan baru. Ini memberikan dukungan penuh untuk proses manajemen dinamis dalam memperoleh pelanggan, mempertahankan pelanggan dan meningkatkan nilai pelanggan, kepuasan pelanggan dan mempromosikan loyalitas pelanggan. Membangun hubungan pemetaan antara atribut konsepsi dan pelanggan adalah langkah kunci dari metode segmentasi berdasarkan data mining [18]. Data pelanggan mengandung atribut dispersive dan continue. Menetapkan setiap atribut pelanggan sebagai dimensi dan menetapkan setiap pelanggan sebagai partikel, seluruh pelanggan di perusahaan dapat membentuk ruang multidimensi, yang telah didefinisikan sebagai ruang atribut pelanggan. Hubungan pemetaan antara atribut pelanggan dan kategori konsepsi dapat dibangun dengan metode analitik, atau dengan metode pembelajaran sampel. Metode analitik menganalisis karakter atribut dari setiap kategori konsepsi yang harus dimiliki, selanjutnya membangun pemetaan hubungan antara ruang atribut dan ruang konsepsi[12]. Namun banyak hubungan pemetaan antara ruang atribut dan ruang konsepsi yang tidak jelas, maka perlu menggunakan metode pembelajaran sampel untuk membangun hubungan pemetaan [11]. Metode pembelajaran sampel secara otomatis menggeneralisasikan hubungan pemetaan antara ruang atribut dan ruang konsepsi dengan menerapkan teknologi data mining pada kategori konsepsi yang diketahui dalam database perusahaan. Proses data mining disebut pembelajaran sampel.Adapun aturan dalm membuat klasifikasi pelanggan[10].

1) Pembuatan aturan segmentasi

Urutkan pelanggan dengan Model Segmen Pelanggan dan Fungsi Segmen. Setelah melatih model segmen, kami mendapatkan aturan segmen atau aturan segmen jaringan. Kami dapat secara efektif mengelompokkan pelanggan baru berdasarkan model yang terlatih.

2) Analisis Fungsi

Analisis fungsi meliputi analisis nilai pelanggan, analisis kredit dan analisis promosi, dll, berdasarkan dasar pemetaan hubungan antara pelanggan dan konsep. Selanjutnya, kebutuhan fungsi baru akan dibawa ke CRM dengan berkembangnya praktik manajemen. Fungsi permintaan baru akan ditambahkan ke dimensi konseptual dan merekonstruksi hubungan pemetaan dengan karakteristik pelanggan.

Keuntungan dari membuat segemntasi pelanggan [10]

(1) Meningkatkan efek promosi

Segmentasi pelanggan berdasarkan data mining dapat membantu perusahaan untuk membuat strategi promosi yang sesuai, pada waktu yang tepat, dengan produk dan layanan yang sesuai, yang ditujukan untuk pelanggan yang sesuai.

(2) Menganalisis nilai pelanggan dan loyalitas pelanggan Nilai pelanggan dan loyalitas pelanggan penting untuk strategi dan taktik manajemen perusahaan. Perusahaan dapat mengkonfirmasi peringkat pelanggan sesuai dengan nilai yang diharapkan dan loyalitas yang dianalisis dengan model segmentasi berdasarkan data mining.

(3) Menganalisis risiko kredit

Penilaian risiko adalah cara yang efektif untuk mengevaluasi jenis risiko pelanggan tertentu, biasanya risiko gagal bayar.

(4) Menginstruksikan R&D produk baru

Perusahaan dapat mengetahui preferensi pelanggan mereka dengan menganalisis pelanggan berdasarkan data mining, dan memastikan bahwa berbagai permintaan akan diwujudkan dalam desain baru.

(5) Mengonfirmasi target pasar

Segmentasi pelanggan berdasarkan data mining dapat membuat target kelompok pelanggan jelas dan menemukan pasar secara eksplisit.

Peran kunci pemasaran adalah mengidentifikasi pelanggan atau segmen dengan potensi penciptaan nilai terbesar dan menargetkan mereka dengan sukses dengan strategi pemasaran yang sesuai untuk mengurangi risiko pelanggan bernilai seumur hidup tinggi ini membelot ke pesaing [10]. Dalam mode konstruksi ini, segmentasi pelanggan adalah pekerjaan dasar penambangan data menurut informasi segmentasi historis yang diketahui[12]. Data pelatihan yang digunakan untuk membangun mode perkiraan segmen dapat berupa data historis atau data eksogen yang diperoleh dari pengalaman atau survei. Karena perilaku pelanggan tidak pasti dan tidak konsisten, peneliti dan manajer harus membangun model segmentasi pelanggan yang dinamis untuk mencerminkan karakteristik secara objektif [13]. Di era customer-centric, hasil segmentasi pelanggan berkaitan dengan penetapan strategi dan taktik perusahaan. Praktik terbaik menuntut pemasar mengembangkan pemahaman mereka tentang segmentasi pelanggan berdasarkan teknik penambangan data dan menggunakan output untuk mengembangkan strategi pemasaran secara kreatif untuk memaksimalkan nilai pemegang saham [14].

2.2 Machine Learning in Customer Relationship Management (CRM) Framework

Pembelajaran mesin dan penambangan data membantu perusahaan membuat alat yang dapat membuat dan mengambil tindakan berdasarkan pengetahuan dan informasi pelanggan. Informasi pelanggan adalah dasar untuk menjaga hubungan jangka panjang dengan pelanggan dan juga dikenal sebagai hubungan dan manajemen pelanggan (CRM) [3]. Klasifikasi dan segmentasi kumpulan data pelanggan digunakan untuk menjaga hubungan yang efisien dengan pelanggan dan selanjutnya meningkatkan profitabilitas dan produktivitas. Dalam makalah ini, penulis mengusulkan segmentasi pelanggan berdasarkan properti demografis seperti jenis kelamin, usia dan skor pengeluaran dan menganalisis kumpulan data untuk fakta menarik. Kumpulan data atribut turunan diselidiki untuk klasifikasi. Klasifikasi digunakan untuk mengkategorikan setiap pelanggan ke dalam beberapa kelas, yaitu, 'emas', 'perak', 'elit' dan 'sesekali'. Perbandingan algoritma klasifikasi yang berbeda disimulasikan dengan tool WEKA. Multi-layer perceptron (MLP) ditemukan sebagai algoritma klasifikasi terbaik dengan akurasi 98,33% dibandingkan dengan Naïve Bayes, regresi dan J48 [4].

Revolusi digital dan peningkatan jumlah data yang dihasilkan oleh perusahaan/organisasi dalam beberapa dekade terakhir telah menyebabkan minat yang besar di bidang pembelajaran mesin dan pembelajaran mendalam [4]. Organisasi dan perusahaan teknologi menggunakan analitik prediktif berbasis Machine Learning untuk mendapatkan keunggulan atas pesaing mereka. Tujuan teknik ML adalah untuk menemukan informasi 'tersembunyi' dalam data, yang hampir tidak mungkin dilakukan dengan cara tradisional berdasarkan keterampilan analitis manusia [5]. Menurut (Cioca et al., 2013; Rahman dan Khan, 2017). Teknik Machine Learning digunakan untuk menambang data untuk intelijen bisnis dan strategi yang dapat dijual untuk pelanggan seperti klasifikasi mereka dalam kategori yang berbeda, membuat strategi skema promosi, dan untuk meningkatkan manajemen hubungan pelanggan (CRM) Dalam skenario saat ini, proses bisnis menjadi semakin berorientasi pada pelanggan dan ditempatkan sebagai prioritas utama manajemen [6]. Karena kemajuan teknologi dalam e-commerce, M-commerce, pemasaran virtual dan pemasaran digital, setiap produk hanya berjarak satu klik dari pelanggan. Berdasarkan pernyataan dari (Singh dan Agrawal, 2019a; Adebiyi et al., 2016) Hal ini menyebabkan persaingan yang sangat ketat, yang diperlukan untuk memastikan bahwa konsumen menerima standar kualitas setinggi mungkin untuk mempertahankannya dan mengurangi tingkat pengadukan [7]. Fokus CRM adalah untuk memperluas layanan pelanggan dan dukungan dalam retensi pelanggan. Pelanggan sangat penting bagi setiap perusahaan dan organisasi. Untuk identifikasi dan retensi pelanggan target mereka, sangat penting untuk memiliki analisis data, yang digunakan untuk mengeksplorasi wawasan dan tren yang berharga untuk mengetahui metrik dan sifat pelanggan. Menurut dari (Singh et al., 2018; Yadav et al., 2018 pentingnya untuk mengamati faktor-faktor kunci terpenting yang mempengaruhi keputusan pembelian pelanggan untuk membeli produk dan layanan apa pun [8]. Machine Learning adalah salah satu analisis data populer yang mengatur struktur model analitis, yang berharga untuk pertumbuhan dalam perilaku pembelian. Teknik ML banyak digunakan dalam prediksi segmentasi pelanggan, nilai umur pelanggan (CLTV), churning, penjualan, dll. Berdasarkan pernyataan dari (Sgaier et al., 2017) Segmentasi pelanggan berguna dalam memahami sub populasi demografis dan psikografis apa yang ada di dalam pelanggan Anda dalam kasus bisnis dan memanfaatkan informasi ini untuk meningkatkan laba, citra, nilai, dan manajemen inventaris [9]. Perusahaan di bisnis mana pun sudah menyadari bahwa mendapatkan pelanggan baru tidak cukup untuk kesuksesan yang langgeng dan upaya perlu dilakukan untuk mengidentifikasi segmentasi pelanggan menuju retensi. Dalam makalah ini, teknik segmentasi dan klasifikasi pelanggan untuk tujuan analisis intelijen bisnis diusulkan[10]. Properti demografis pelanggan diambil sebagai parameter segmentasi pelanggan untuk mengetahui analitik tentang pelanggan. Itu akan membantu dalam CRM, efisiensi dan produktivitas pusat perbelanjaan. Selanjutnya, teknik pembelajaran mesin digunakan untuk memprediksi penilaian pelanggan [9].

2.2.1 Customer Relationship Management using Logistic Regression

Pemeriksaan kekambuhan adalah jenis metode menampilkan prescient yang mengeksplorasi hubungan antara yang membutuhkan (target) dan variabel otonom [11]. Prosedur ini digunakan untuk mengantisipasi, menunjukkan pengaturan waktu dan menemukan hubungan sebab akibat antara faktor-faktor tersebut[12]. Misalnya, hubungan antara mengemudi terburu-buru dan jumlah kecelakaan di jalan oleh pengemudi paling baik terkonsentrasi melalui kekambuhan. Regresi Strategis adalah perhitungan pengelompokan Machine Learning yang digunakan untuk meramalkan kemungkinan variabel yang jelas[10]. Dalam relaps strategis, variabel terikat adalah variabel berpasangan yang berisi informasi berkode 1 (benar-benar, prestasi, dan sebagainya.) atau 0 (tidak, kecewa, dan sebagainya.). Model relaps strategis memprediksi P(Y=1) sebagai komponen X. Regresi Terhitung adalah salah satu pendekatan paling terkenal untuk menyesuaikan model untuk informasi yang jelas, terutama untuk informasi reaksi ganda dalam pemodelan data [12].

2.2.2 Customer Relationship Management using Decision Tree

Pohon keputusan dikatakan sebagai salah satu metodologi umum, yang digunakan dalam memprediksi serta memperkirakan masalah churn pelanggan. Berdasarkan metode membagi dan menaklukkan, keputusan dikembangkan[13]. Namun pohon keputusan memang memiliki beberapa batasan seperti tidak dapat digunakan untuk hubungan non-linier dan kompleks antar atribut. Namun, telah diamati bahwa metode pohon keputusan memang mencoba untuk meningkatkan akurasi klasifikasi [13]. Dalam makalah JST bersama dengan pohon keputusan digunakan untuk melakukan prediksi churn pelanggan dan diketahui bahwa pohon keputusan mengungguli jaringan saraf dalam hal akurasi[14]. Mempresentasikan metodologi klasifikasi aplikasi DT untuk menganalisis churn rate di industri telekomunikasi. Di sini, pohon keputusan ID3 digunakan dan diamati bahwa area pelanggan adalah salah satu fitur klasifikasi utama, yang lain memberikan dua hasil pada pelanggan untuk churn menggunakan beberapa metode seperti K-means clustering, QUEST, CART, Logistic Regression, neural network, exhaustive CHAId [13]. Di sini, diamati bahwa CHAID berkinerja jauh lebih baik daripada metodologi lain yang disebutkan. Diperhatikan bahwa akurasinya sekitar 60%, yang jauh lebih baik daripada metodologi lainnya. Selain itu, pohon keputusan lainnya tidak berdiri di depan untuk Exhaustive CHAID [14].

2.2.3 Customer Relationship Management using Random Forest Tree

Hutan acak adalah algoritma pembelajaran terawasi yang digunakan baik untuk klasifikasi maupun regresi [15]. Tapi bagaimanapun, ini terutama digunakan untuk masalah klasifikasi. Seperti yang kita ketahui bahwa hutan terdiri dari pohon dan lebih banyak pohon berarti hutan yang lebih kuat. Demikian pula, algoritma hutan acak membuat pohon keputusan pada sampel data dan kemudian mendapatkan prediksi dari masing-masing dan akhirnya memilih solusi terbaik melalui pemungutan suara [16]. Ini adalah metode ensemble yang lebih baik daripada pohon keputusan tunggal karena mengurangi over-fitting dengan merata-ratakan hasilnya. Ini mengatasi masalah overfitting dengan merata-ratakan atau menggabungkan hasil dari pohon keputusan yang berbeda. Hutan acak bekerja dengan baik untuk sejumlah besar item data daripada pohon keputusan tunggal. Hutan acak memiliki varians yang lebih sedikit daripada pohon keputusan tunggal. Hutan acak sangat fleksibel dan memiliki akurasi yang sangat tinggi [10]. Penskalaan data tidak diperlukan dalam algoritma hutan acak. Ini mempertahankan akurasi yang baik bahkan setelah memberikan data tanpa penskalaan. Algoritma Random Forest mempertahankan akurasi yang baik bahkan sebagian besar data hilang[15]. Kompleksitas adalah kelemahan utama dari algoritma hutan acak. Konstruksi hutan acak jauh lebih sulit dan memakan waktu daripada pohon keputusan. Lebih banyak sumber daya komputasi diperlukan untuk mengimplementasikan algoritma Random Forest. Ini kurang intuitif jika kita memiliki banyak koleksi pohon keputusan [17]. Proses prediksi menggunakan hutan acak sangat memakan waktu dibandingkan dengan algoritma lain. Beberapa algoritma konvensional seperti Pohon Keputusan, algoritma Genetika, jaringan saraf dan pohon klasifikasi telah diusulkan. Algoritma tersebut di atas mampu memperkirakan tingkat churn. Namun, mereka memiliki beberapa masalah seperti pohon keputusan yang kurang dengan masalah probabilitas kelas yang sama yang secara drastis dapat mengurangi kinerja [18]. Demikian pula, dalam kasus algoritma genetika, sangat tidak mungkin untuk mengenali kemungkinan terkait dengan estimasi yang menghasilkan kinerja rendah dan dalam hal teknik seperti state-of-art menghasilkan beberapa kesalahan. Oleh karena itu, pembahasan di atas dapat menyimpulkan bahwa teknik RF konvensional tidak menghasilkan hasil yang efisien dalam hal kumpulan data yang besar dan juga berkinerja sangat buruk ketika mereka tidak seimbang. Metode MRF (Modified Random Forest) yang kami usulkan berkinerja lebih baik di berbagai parameter seperti ketahanan akurasi dan lainnya [19]. Skema kami terutama memiliki lapisan tambahan RV (Variabel Acak) yang membantu untuk melakukan model jauh lebih baik bersama dengan metode kami membantu dalam meminimalkan kebisingan Gaussian dan juga membantu dalam mengurangi masalah regresi dan klasifikasi. Metode kami membantu dalam membangun berbagai pohon yang berbeda dari set data pelatihan tertentu. Metode yang diusulkan membantu dalam memfokuskan estimasi tingkat churn konsumen dalam layanan telekomunikasi. Pada penelitian yang dilakukan oleh Irfan dkk Algoritma Random Forest (RF) bekerja dengan baik dengan 88,63% instance yang diklasifikasikan dengan benar[20]. Membuat kebijakan retensi yang efektif adalah tugas penting dari CRM untuk mencegah churner. Setelah klasifikasi, model yang diusulkan menyegmentasikan data pelanggan yang berpindah dengan mengkategorikan pelanggan yang berhenti dalam kelompok menggunakan kesamaan kosinus untuk memberikan penawaran retensi berbasis grup. Makalah ini juga mengidentifikasi faktor-faktor churn yang penting dalam menentukan akar penyebab churn [21]. Dengan mengetahui faktor-faktor churn yang signifikan dari data pelanggan, CRM dapat meningkatkan produktivitas, merekomendasikan promosi yang relevan kepada kelompok pelanggan yang kemungkinan churn berdasarkan pola perilaku yang serupa, dan secara berlebihan meningkatkan kampanye pemasaran perusahaan. Model prediksi churn yang diusulkan dievaluasi menggunakan metrik, seperti akurasi, presisi, recall, f-measure, dan area karakteristik operasi penerima (ROC)[22]. Hasilnya mengungkapkan bahwa model prediksi churn yang kami usulkan menghasilkan klasifikasi churn yang lebih baik menggunakan algoritma RF dan profil pelanggan menggunakan k-means clustering. Selain itu, juga memberikan faktor-faktor di balik churn pelanggan churn melalui aturan yang dihasilkan dengan menggunakan algoritma pengklasifikasi yang dipilih atribut [17].