CUSTOMER SEGMENTATION IN CUSTOMER RELATIONSHIP MANAGEMENT BASED ON DATA MINING by yung chen

Makalah ini akan membangun pelanggan model fungsi segmentasi berdasarkan data mining, dan merangkum keuntungan dari model fungsi segmentasi pelanggan berdasarkan data mining di manajemen hubungan pelanggan (CRM).

Selama dekade terakhir, telah terjadi ledakan minat dalam manajemen hubungan pelanggan (CRM) oleh akademisi dan eksekutif [1]. Organisasi menyadari bahwa pelanggan memiliki nilai ekonomi yang berbeda bagi perusahaan, dan mereka kemudian mengadaptasi penawaran pelanggan dan strategi komunikasi yang sesuai. Dengan demikian, organisasi, pada dasarnya, bergerak menjauh dari pemasaran produk atau merek-sentris menuju pendekatan yang berpusat pada pelanggan. Saat ini penelitian menunjukkan bahwa penerapan aktivitas CRM menghasilkan kinerja perusahaan yang lebih baik ketika manajer fokus untuk memaksimalkan nilai pelanggan [2]. Segmentasi pelanggan adalah dasar bagaimana memaksimalkan nilai pelanggan. Lagi dan lagi perusahaan menemukan bahwa prinsip Pareto berlaku, dengan 20% dari basis pelanggan menghasilkan 80% dari keuntungan. Baik peneliti maupun manajer perlu mengevaluasi dan memilih segmentasi untuk merancang dan menetapkan strategi yang berbeda untuk memaksimalkan nilai pelanggan.

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.

Metode statistik tradisional mengelompokkan pelanggan menurut karakter perilaku sederhana atau karakter atribut seperti kategori produk yang dibeli atau wilayah tempat tinggalnya.

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. 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. 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.

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

(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. 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. 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.

Review of Data Mining Techniques for Churn Prediction in Telecom by by vishal Mahajan

Churn : memutuskan hubungan dengan pelanggan

Sektor telekomunikasi menghasilkan sejumlah besar data karena meningkatnya jumlah pelanggan, teknologi terbarukan dengan cepat; aplikasi berbasis data dan layanan nilai tambah lainnya. Data ini dapat ditambang dengan berguna untuk analisis dan prediksi churn. Penelitian signifikan telah dilakukan oleh para peneliti di seluruh dunia untuk memahami praktik penambangan data yang dapat digunakan untuk memprediksi churn pelanggan. Makalah ini memberikan ulasan tentang sekitar 100 artikel jurnal terbaru mulai dari tahun 2000 untuk menyajikan berbagai teknik penambangan data yang digunakan dalam beberapa model churn berbasis pelanggan. Ini kemudian merangkum literatur telekomunikasi yang ada dengan menyoroti ukuran sampel yang digunakan, variabel churn yang digunakan dan temuan teknik DM yang berbeda. Akhirnya, kami mencantumkan teknik paling populer untuk prediksi churn di telekomunikasi sebagai pohon keputusan, analisis regresi, dan pengelompokan, sehingga memberikan peta jalan bagi peneliti baru untuk membangun model manajemen churn baru.

Pelanggan saat ini memiliki persediaan informasi yang tampaknya tak ada habisnya di ujung jari mereka Akibatnya, perusahaan di berbagai industri mengalami kesulitan dalam menarik dan mempertahankan pelanggan. Karena kemajuan teknologi yang pesat dan persaingan yang semakin ketat, pelanggan memiliki banyak pilihan untuk dipilih dan ini telah menjadi tantangan bagi telekomunikasi operator. Perusahaan kehilangan banyak pendapatan karena beralih oleh pelanggan mereka yang sudah ada.

Churn bisa bersifat sukarela dan tidak sukarela. churn sukarela terjadi ketika pelanggan yang sudah ada meninggalkan penyedia layanan dan bergabung dengan penyedia layanan lain, sedangkan dalam penghentian paksa pelanggan diminta oleh penyedia layanan untuk pergi karena alasan seperti tidak membayar, dll. [34],[57]. churn sukarela dapat dibagi menjadi: churn insidental dan churn disengaja [35]. Churn insidental terjadi, bukan karena pelanggan merencanakannya tetapi karena sesuatu terjadi dalam hidup mereka mis. perubahan kondisi keuangan, perubahan lokasi, dll. Perpindahan yang disengaja terjadi karena alasan teknologi (pelanggan menginginkan teknologi yang lebih baru atau lebih baik, sensitivitas harga, faktor kualitas layanan, faktor sosial atau psikologis, dan alasan kenyamanan) [69]. Dia juga menekankan untuk memprediksi bagaimana setiap pelanggan akan bereaksi terhadap penawaran tertentu dan memprediksi pelanggan mana yang akan terpengaruh secara positif.

jika pelanggan yang ada memutuskan kontrak dengan satu penyedia layanan dan menjadi pelanggan dari penyedia layanan lain, maka pelanggan ini disebut sebagai pelanggan 'hilang' atau pelanggan 'Churn'.

Penambangan data memprediksi tren dan perilaku masa depan, yang membantu bisnis menjadi lebih proaktif dan memungkinkan mereka mengambil keputusan berdasarkan pengetahuan. Data mining dapat menjawab pertanyaan bisnis yang secara tradisional terlalu memakan waktu untuk diselesaikan.

Question

Apa saja area aplikasi berbasis pelanggan spesifik yang metode DM telah diterapkan?

• Apa metode DM yang umum digunakan yang diterapkan di domain ini?

• Pada dimensi churn dan ukuran sampel seperti apa, apakah metode tersebut beroperasi di sektor telekomunikasi?

• Untuk meringkas teknik DM paling populer yang digunakan dalam model prediksi churn di sektor telekomunikasi.

• Untuk menunjukkan daerah subur untuk pekerjaan penelitian lebih lanjut di lapangan.

Tujuan dari data mining (DM) adalah untuk menganalisis kumpulan data yang besar untuk mendapatkan kembali yang bermakna informasi. untuk memprediksi perilaku pelanggan yang kemungkinan besar akan keluar dari layanan penyedia layanan yang ada dan bergabung dengan penyedia layanan baru. Memahami tren saat ini dan masa lalu, perilaku dan perencanaan untuk masa depan adalah penting dalam bisnis. Oleh karena itu, aplikasi penambangan data memainkan peran penting dalam pengambilan keputusan dan memberikan prediksi perkiraan masa depan. Fungsionalitas data mining utama dapat diklasifikasikan sebagai berikut: multivariat analisis statistik (analisis regresi), penambangan hubungan (penambangan pola sering algoritma), pengelompokan, klasifikasi (pohon keputusan, jaringan saraf), prediksi dan outlier deteksi [10], [38].

Pemodelan prediktif pada dasarnya berkaitan dengan meramalkan bagaimana pelanggan akan berperilaku di masa depan dengan menyelidiki perilaku masa lalu mereka [40]. Mengantisipasi pelanggan yang cenderung churn adalah salah satu contoh pemodelan prediktif. Ini digunakan dalam menganalisis Pelanggan Data Manajemen Hubungan (CRM) dan penambangan data (DM) untuk memberikan layanan berbasis pelanggan model yang menggambarkan probabilitas bahwa pelanggan akan mengambil tindakan tertentu [73]. Ini tindakan dapat berupa penjualan, pemasaran, dan churn/retensi pelanggan terkait. Ada banyak model yang dapat digunakan untuk membedakan antara churners dan non-churners dalam suatu organisasi. Peneliti [74] mengklasifikasikan dimensi Manajemen Hubungan Pelanggan menjadi empat set yaitu Identifikasi Pelanggan, Daya Tarik Pelanggan, Retensi Pelanggan, dan Pelanggan Pengembangan menggunakan fungsi data mining populer seperti Asosiasi, Klasifikasi, Clustering, Forecasting, Regression, Sequence Discovery dan Visualization.

untuk memprediksi churner di industri telekomunikasi Taiwan, yang telah menjadi fokus utama industri. Mereka menganggap demografi pelanggan (usia, masa kerja, jenis kelamin), informasi penagihan dan pembayaran (jumlah tagihan, biaya bulanan, jatuh tempo pembayaran), detail panggilan (durasi panggilan, jenis panggilan), layanan layanan pelanggan dari pelanggan dari salah satu perusahaan telekomunikasi termasuk churners untuk jangka waktu satu tahun. Mereka tersegmentasi pelanggan ke dalam berbagai cluster menggunakan K-means clustering berdasarkan jumlah, tenurial, penggunaan panggilan keluar, penggunaan panggilan masuk, dan tarif pembayaran. Penulis menemukan bahwa perusahaan pengguna memiliki kemungkinan tinggi untuk churn, mungkin karena perubahan pekerjaan. Mereka menyarankan agar pengguna yang tidak melakukan panggilan ke pengguna lain di jaringan yang sama memiliki kemungkinan besar untuk melakukan churn. Mereka juga menemukan bahwa pengguna yang kontraknya akan berakhir dalam waktu dekat memiliki lebih banyak kemungkinan untuk bergejolak. Penulis [93] melakukan penelitian di Turki, di mana sektor telekomunikasi sangat menderita kehilangan pelanggan, untuk mengetahui jenis pelanggan apa yang beralih penyedia layanan dan apa adalah alasan di balik itu. Penulis menganalisis catatan 1000 pelanggan untuk jangka waktu 6 bulan dan menggunakan teknik penambangan Regresi Logistik dan Pohon Keputusan. Mereka menggunakan penggunaan pelanggan sebagai parameter untuk memprediksi churn dan menemukan bahwa jika pelanggan tidak memiliki paket diskon apa pun, maka ada kemungkinan 75% pelanggan akan churn. Penulis juga menemukan bahwa faktor penting lainnya yang bertanggung jawab untuk churn adalah jumlah panggilan masuk dan lama panggilan jarak jauh. Pelanggan yang menerima panggilan maksimum dari pelanggan yang menggunakan layanan yang sama penyedia layanan cenderung churn dibandingkan pelanggan yang menggunakan penyedia layanan lain.

Conclusion

Temuan survei menunjukkan berbagai teknik DM yang digunakan untuk memprediksi churn pelanggan dari tahun 2000 dan 2014 di domain yang berbeda mis. perbankan, surat kabar (media), ritel dan analisis risiko kredit. Ini menguraikan tren saat ini, dimensi yang digunakan dan tantangan aplikasi DM di telekomunikasi, bagi para peneliti yang mulai meninjau bidang ini.

Berdasarkan tinjauan literatur, berbagai teknik penambangan yang digunakan untuk churn sejauh ini, diklasifikasikan pada Tabel 1 di atas. Teknik DM yang paling populer adalah pohon keputusan, regresi, jaringan saraf

dan clustering seperti yang digambarkan pada Gambar 1 di atas. Namun, tidak ada konsensus umum yang jelas tentang teknik prediksi yang akan digunakan pada data yang dikumpulkan [100]. Selanjutnya, karena biaya yang terlibat, sebagian besar studi yang ada yang terlibat dalam survei menggunakan sampel data kecil dari catatan pelanggan [52], [12], [33], [66], [59], yang dapat merusak keandalan dan validitas hasil analisis. Ini berarti studi empiris dengan kumpulan data signifikan yang lebih besar dengan dimensi tambahan dapat meningkatkan keandalan hasil. Disarankan agar pengembangan teknik DM ke depan dapat menjadi lebih berorientasi pada masalah

dan khusus untuk prediksi 'tipe churner' diperlukan. Selain itu, model hybrid dapat diperkenalkan dan dibandingkan dengan model yang ada. Ini akan membantu dalam merancang kumpulan data pelanggan multidimensi dan merancang teknik manajemen churn baru yang spesifik untuk kumpulan data yang berbeda dan lokasi geografis yang berbeda. Pengambilan keputusan berdasarkan analisis dari teknik DM dapat membuat prediksi churn lebih akurat dan akan memberikan wawasan yang berharga untuk teknik industri seluler.

Market Segmentation, Targeting, and Positioning by Syed Saad Andaleeb

Strategi sering diarahkan pada kelompok pelanggan terbesar (misalnya, dalam kota besar, di terminal bus atau kereta api, di jalan-jalan yang sibuk, tempat pertemuan keagamaan, pasar desa besar, dll) dengan menawarkan standar, barang murah yang konon akan dijual karena fungsi produk dan harga murah. Strategi seperti itu mengabaikan lebih dalam karakteristik pelanggan, selera yang bervariasi, atau keinginan mereka untuk berdiri keluar dan menjadi berbeda. Dengan menyesuaikan produk agar sesuai dengan demografi, psikografis, geografis, atau karakteristik lain yang berbeda, atau dengan berfokus pada kebutuhan atau keinginan pelanggan tertentu, dan mempertimbangkan bagaimana mereka cenderung merespons secara perilaku terhadap penawaran tertentu, daya tarik dan keinginan produk dapat secara signifikan ditingkatkan.

Kustomisasi semacam itu mewakili fokus pada pasar yang tersegmentasi terdiri dari sekumpulan pelanggan yang memiliki kebutuhan yang sama dan menginginkan serta berperilaku serupa dengan daya tarik dan penawaran pemasaran. Kumpulan pelanggan ini disebut segmen pasar. tindakan dari segmentasi pasar, oleh karena itu, merupakan upaya untuk mengidentifikasi dan mengkategorikan kelompok pelanggan ke dalam kelompok yang menunjukkan beberapa kesamaan sifat. Pengelompokan juga dapat terjadi di makro tingkat dengan mengelompokkan komunitas, wilayah, atau bahkan negara menurut beberapa karakteristik umum (misalnya, bahasa, adat istiadat, agama,

Segmentasi didefinisikan sebagai proses mengidentifikasi kelompok pelanggan potensial tertentu dengan atribut yang homogen yang cenderung menunjukkan tanggapan serupa terhadap penawaran atau bauran pemasaran. Karena semakin banyak pemasar mulai menyadari bahwa ada tradeoff antara proposisi harga-nilai standar dan eksklusivitas-kekhususan etos yang mungkin diinginkan oleh pelanggan, pilihan strategi yang lebih efektif mulai terbuka: pemasaran mikro. Pemasaran mikro memungkinkan pemasar untuk membuat penawaran dengan menyempurnakan program pemasarannya pada serangkaian atribut yang terdefinisi dengan baik sebagai penawaran kompetitifnya. Kuncinya adalah menentukan level di mana pengelompokan dilakukan (untuk organisasi berita: umum berita, berita olahraga, atau hanya berita kriket) di mana penawarannya berbeda dan kekhasan itu dapat dilindungi untuk mendominasi suatu segmen pasar tertentu (yaitu, khalayak umum, khalayak olahraga, atau penonton kriket).

Tantangan yang dihadapi pemasar mikro ada tiga:

• Bagaimana mengidentifikasi kelompok pelanggan potensial yang serupa (segmentasi)?

Ini melibatkan proses mendefinisikan segmen yang memiliki potensi yang signifikan untuk menanggapi upaya pemasaran.

Bagaimana memilih segmen yang akan ditargetkan (penargetan)?

Pemasar harus memutuskan segmen yang dapat dijangkau paling efektif, efisien, jelas, dan menguntungkan.

• Bagaimana memposisikan produk/merek di benak pelanggan (posisi)?

Positioning diperlukan untuk membedakan produk atau merek di benak pasar sasaran, menggambarkannya sebagai penawaran pada atribut yang dipilih.

Segmentasi pasar

Tidak ada cara tunggal untuk mengidentifikasi segmen pasar. Pemasar harus selalu mengevaluasi hubungan produk-pasar; yaitu, untuk kelompok pelanggan mana yang akan ditawarkan?

dibuat. Pengelompokan tersebut mungkin atas dasar geografi, demografi, psikografis, perilaku, dan manfaat. Sementara pemahaman konseptual dan intuitif pelanggan di berbagai segmen itu penting, segmentasi yang efektif umumnya didukung oleh inisiatif berbasis data. Dengan demikian, pengumpulan dan analisis data dapat memainkan peran penting dalam latihan segmentasi pasar.

Penargetan Pasar

Setelah perusahaan mengembangkan peluang segmen pasarnya,

harus memutuskan segmen atau segmen mana yang akan diprioritaskan dan dipelihara,

dan mana yang harus diabaikan atau ditetapkan sebagai prioritas rendah. Tiga kriteria dasar

harus dipenuhi untuk mengasah target yang dipilih. Ini

antara lain sebagai berikut:

• Ukuran segmen saat ini dan potensi pertumbuhan yang diantisipasi

• Potensi persaingan di segmen tersebut

• Kompatibilitas dengan tujuan perusahaan secara keseluruhan dan

kelayakan untuk berhasil mencapai target audiens.

Para ahli percaya mungkin lebih baik untuk melihat kecil tapi tumbuh

segmen alih-alih mengincar segmen besar di mana pesaing mungkin telah memantapkan diri mereka; sehingga memberikan rendah

untuk tidak ada peluang pertumbuhan. Ketika kamera digital diperkenalkan, pasar potensial adalah faktor pendorongnya, bukan pasar besar kamera berbasis film yang ada. Dengan demikian, ukuran pasar dan pertumbuhan yang diantisipasi merupakan pertimbangan awal dalam memilih target tertentu. Selain menjalankan program pemasaran yang sukses di segmen tertentu, akan berguna, seperti yang disarankan sebelumnya, jika segmen tersebut dapat diukur (melalui survei dan sarana statistik terkait), substansial (potensi keuntungan), dapat diakses (melalui saluran komunikasi dan distribusi yang berbeda) , dapat dibedakan (memiliki fungsi respons yang berbeda), dapat ditindaklanjuti (melalui program pemasaran yang efektif), dan konsisten dengan tujuan dan sumber daya perusahaan.

Business analytics use in CRM: A nomological net from IT competence to CRM performance by Dalwoo Nam

Analisis bisnis (BA) menjadi semakin penting di bawah lingkungan bisnis yang berubah dengan cepat. Sebuah penelitian tantangannya adalah bahwa penggunaan BA tidak sepenuhnya dipahami. Kami mengatasi tantangan ini dari perspektif dinamis kapabilitas dengan menggunakan model empiris dengan penekanan pada penggunaan BA dalam manajemen hubungan pelangga (CRM). Berdasarkan 170 sampel dari survei tingkat perusahaan, kami menganalisis hubungan nomologis dari kompetensi TI ke kinerja CRM. Hasilnya menunjukkan kemampuan manajemen data sepenuhnya memediasi antara kompetensi TI dan BA digunakan, sementara kemampuan respons pelanggan sebagian menengahi antara penggunaan BA dan kinerja CRM.]

Analytics telah diinvestasikan secara aktif dan diadopsi secara luas untuk mendukung teknologi manajemen hubungan pelanggan (CRM) (yaitu memperoleh dan mempertahankan pelanggan). Vendor dan praktisi CRM percaya bahwa kemampuan CRM yang mendukung BA, kemampuan untuk menganalisis, mengintegrasikan, dan memanfaatkan sumber daya informasi dan umpan balik pelanggan untuk dukungan pengambilan keputusan di CRM, akan meningkatkan nilai bisnis dan aktif dalam memperkenalkan sistem BA dengan pemodelan statistik canggih, simulasi, peramalan dan pembelajaran mesin (Coltman, 2007; Coltman, Devinney, & Midgley, 2011; Shanks & Bekmamedova, 2012). Namun, hasil penggunaan teknologi CRM masih dapat diperdebatkan (misalnya, Reinartz, Krafft, & Hoyer, 2004), dan hanya ada sedikit penelitian tentang penggunaan BA untuk meningkatkan kinerja CRM. Karena CRM secara tradisional merupakan dasar dari analisis data dan sekarang sedang ditransformasikan menjadi berbasis big data atau analisis bisnis, maka perlu untuk mempelajari serangkaian proses yang menghasilkan nilai bisnis dan hasilnya dapat menjadi panduan untuk penggunaan dan penerapan BA untuk praktisi. Dengan demikian, penelitian ini mencoba untuk mengeksplorasi jaring nomologis dengan berfokus pada penggunaan BA dalam mendorong nilai bisnis dari kompetensi TI dalam konteks CRM.

Penelitian ini untuk menjelaskan lebih banyak tentang kemampuan dinamis yang dapat memanfaatkan BA yang mengarah pada kinerja CRM yang lebih baik dari kompetensi TI.

Studi ini bermakna dalam sistem informasi (IS) literatur. Kita mengembangkan model pelit dengan memeriksa anteseden dan konsekuensi dari penggunaan BA. Kami mengumpulkan data dari survei tingkat perusahaan dan 170 sampel digunakan untuk menguji model penelitian kami. Dengan menjelajahi hubungan seputar penggunaan BA, kami menemukan mata rantai yang hilang dari kompetensi TI hingga kinerja bisnis melalui penggunaan BA. Dari perspektif kemampuan dinamis, kami meningkatkan pemahaman jaring nomologis dari kompetensi TI hingga kinerja CRM dengan mendefinisikan kemampuan respons pelanggan dan kemampuan manajemen data. Studi ini dapat menjadi batu loncatan untuk studi masa depan tentang penggunaan BA dalam CRM dan kemampuan dinamis.

A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector by Irfan

Di sektor telekomunikasi, sejumlah besar data dihasilkan setiap hari karena basis klien yang luas. Pengambil keputusan dan analis bisnis menekankan bahwa mendapatkan pelanggan baru lebih mahal daripada mempertahankan yang sudah ada. Analis bisnis dan analisis manajemen hubungan pelanggan (CRM) perlu mengetahui alasan pelanggan churn, serta, pola perilaku dari data pelanggan churn yang ada. Makalah ini mengusulkan model prediksi churn yang menggunakan klasifikasi, serta, teknik clustering untuk mengidentifikasi pelanggan churn dan memberikan faktor-faktor di balik churn pelanggan di sektor telekomunikasi. Seleksi fitur dilakukan dengan menggunakan filter pemeringkatan atribut perolehan informasi dan korelasi. Model yang diusulkan pertama mengklasifikasikan data pelanggan churn menggunakan algoritma klasifikasi, di mana:

Algoritma Random Forest (RF) bekerja dengan baik dengan 88,63% instance yang diklasifikasikan dengan benar. 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. 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). 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.\

Prediksi churn sangat penting di sektor telekomunikasi karena telekomunikasi

operator harus mempertahankan pelanggan berharga mereka dan meningkatkan

administrasi Manajemen Hubungan Pelanggan (CRM) mereka

[5], [6]. Pekerjaan yang paling menantang bagi CRM adalah mempertahankan pelanggan yang sudah ada [7]. Karena pasar yang jenuh dan kompetitif, pelanggan memiliki opsi untuk beralih ke penyedia layanan lain. Perusahaan telekomunikasi telah mengembangkan prosedur untuk mengidentifikasi dan mempertahankan pelanggan mereka karena lebih murah daripada menarik pelanggan baru [5]. Ini karena biaya yang terlibat dalam iklan, tenaga kerja, dan konsesi yang dapat meningkat hingga hampir lima hingga enam kali lipat daripada mempertahankan pelanggan yang sudah ada [3]. Perhatian kecil diperlukan untuk mengidentifikasi pelanggan churn yang ada, yang dapat membantu membalikkan situasi. Persyaratan untuk mempertahankan pelanggan perlu mengembangkan model yang akurat dan berkinerja tinggi untuk mengidentifikasi pelanggan yang berhenti. Model yang diusulkan harus memiliki kemampuan untuk mengidentifikasi pelanggan churn dan kemudian menemukan alasan di balik churn untuk menghindari kehilangan pelanggan dan memberikan langkah-langkah untuk mempertahankan mereka. Selain itu, harus menggunakan teknik untuk memprediksi kapan situasi seperti itu akan muncul di masa depan.

Karena kemajuan terbaru di bidang data besar, ada ada banyak solusi penambangan data dan pembelajaran mesin yang dapat digunakan untuk menganalisis data tersebut. Teknik-teknik ini menganalisis data dan mengidentifikasi alasan di balik perputaran pelanggan. CRM dapat menggunakan teknik ini untuk memaksimalkan keuntungan mereka [2]. Selanjutnya, dapat digunakan untuk merancang strategi retensi untuk mengurangi rasio pelanggan yang akan churn. CRM dapat mencapai tujuan retensi pelanggan perusahaan dengan mengidentifikasi kebutuhan pelanggan yang akurat dengan menggunakan teknik data mining. Data mining melibatkan proses mengidentifikasi perilaku pelanggan churn dari pola yang diambil dari data. Penambangan data dikenal dengan banyak nama berbeda seperti kecerdasan bisnis, pemodelan prediktif, penemuan pengetahuan, dan analitik prediktif. Penambangan data adalah salah satu dimensi CRM, di mana penganalisis CRM harus menggabungkan tiga dimensi utama; data, penambangan data, dan pengambil keputusan.

CRM membutuhkan konotasi untuk mengenali dan memahami unit bisnis dan pelanggannya. CRM juga mengontrol peningkatan dalam penawaran dan diskon seperti item apa yang ditawarkan kepada pelanggan mana dan layanan serta promosi mana yang mereka membutuhkan.

Customer Churn Prediction and Upselling using MRF (Modified Random Forest) technique by SwethaP

Prediksi Churn Pelanggan telah menjadi salah satu topik terkemuka di industri telekomunikasi, telah mendapatkan banyak perhatian di industri penelitian karena persaingan yang ketat dari berbagai, dan karenanya perusahaan telah berfokus pada ukuran data yang lebih besar untuk pengadukan dan penjualan. ramalan. Model prediksi churn pelanggan mendeteksi dan mengidentifikasi pelanggan yang bersedia menghentikan langganan, prediksi churn pelanggan dan upselling dapat dilakukan melalui proses data mining. Oleh karena itu, Dalam makalah ini kami telah memperkenalkan model Bernama MRF (Modified Random Forest), model ini membantu dalam meningkatkan akurasi dan juga membantu dalam mengabaikan masalah regresi. Metodologi kami telah dilakukan pada Dataset oranye yang disediakan. Untuk evaluasi analisis perbandingan algoritma kami antara metodologi yang ada dan yang diusulkan dilakukan dengan mempertimbangkan dua skenario yaitu churn dan upselling. Kemudian model kami dibandingkan dengan berbagai model prediksi churn yang ada, hasil analisis menunjukkan bahwa model kami mengungguli metode yang ada termasuk hutan acak standar dalam hal AUC dan akurasi klasifikasi.

Dalam beberapa tahun terakhir, telah terjadi generasi yang sangat besar. data; namun memiliki jumlah data yang besar membuatnya lebih penting untuk mengekstrak informasi yang tersembunyi di atas data mentah yang diberikan. Dari sejumlah besar data ini, data yang bermakna dapat diekstraksi untuk membantu industri tumbuh. Salah satu metode yang digunakan untuk mengekstrak informasi dikenal sebagai data mining[3]. Penambangan data tidak lain adalah proses menemukan pola dalam data besar yang diberikan untuk mengidentifikasi pola serta mencoba membangun hubungan untuk menyelesaikan masalah analitik data. Dalam dua dekade terakhir, pertumbuhan volume data telah diperhatikan karena improvisasi hari demi hari di TI (Teknologi Informasi). Secara paralel penelitian mendalam telah terjadi di data mining. Beberapa metode serta teknik baru ditambahkan untuk mengumpulkan informasi yang diperlukan dan memproses data [4]. Data, yang dikumpulkan dari berbagai sumber, adalah data mentah di mana banyak informasi disembunyikan.

Oleh karena itu, dalam penelitian ini kami telah mengusulkan sebuah model yaitu MRF(Modified Random forest) , metode kami menyediakan beberapa lapisan untuk meningkatkan akurasi dan mengabaikan overfitting. Model ini membantu dalam membuang masalah regresi, masalah regresi adalah salah satu masalah biasa-biasa saja yang muncul saat prediksi churn dan upsell. Model kami memberikan tingkat kesalahan minimum dan tingkat klasifikasi yang lebih baik daripada metode canggih lainnya termasuk metode yang ada (Modified Random Forest).

Di masa lalu telah banyak penelitian tentang prediksi churn pelanggan, di bagian ini, kita telah membahas berbagai metodologi yang ada. Survei ini dibagi berdasarkan berbagai metode seperti Neural Network, Pohon keputusan dan algoritma penutup, studi singkat tentang pekerjaan yang ada telah mengarahkan kami untuk merancang model kami yaitu:

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. 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. Dalam makalah [13] 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.[7] menggunakan beberapa metode seperti K-means clustering, QUEST, CART, Logistic Regression, neural network, exhaustive CHAId. 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.

Modified Random Forest Preliminaries

Pada bagian ini, teknik RF (Random Forest) Modifikasi yang diusulkan digunakan. 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. 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. 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.

Churn Prediction using Machine Learning-An Analytical CRM Application by M Kavitha Margret

CRM mewakili (Customer Relationship Management). Ini adalah klasifikasi program yang mencakup banyak pengaturan penggunaan yang dimaksudkan untuk mendukung organisasi dan juga untuk mengawasi sejumlah besar bentuk bisnis seperti informasi klien. Model kerangka kerja CRM menggabungkan tahapan yang bekerja untuk mengawasi periklanan, kesepakatan, dukungan klien, dan dukungan, semua yang terkait dengan organisasi bantuan bekerja dengan lebih layak. Dengan kerangka kerja CRM, organisasi dapat membedah kolaborasi klien dan meningkatkan koneksi klien mereka. Model perkiraan berbasis data yang menggunakan sistem AI telah meningkatkan prevalensi mengerikan selama beberapa dekade terakhir. Model-model ini telah diterapkan di sejumlah besar bidang seperti kesimpulan klinis, harapan kesalahan, peringkat film, dan sebagainya. Oleh karena itu digunakan dalam industri telekomunikasi di mana model harapan telah diterapkan untuk perkiraan klien yang tidak terpenuhi yang mungkin akan mengubah administrasi dan juga organisasi spesialis. Dalam telekomunikasi, biaya terkait transaksi klien sangat besar dan selanjutnya banyak organisasi telah memeriksa variabel yang berbeda, (misalnya, biaya panggilan, sifat panggilan, waktu reaksi bantuan klien, dan sebagainya.) menggunakan strategi AI yang berbeda. Pekerjaan ini mengusulkan strategi ML yang berbeda untuk harapan agitasi klien.

Disini perhitungan Random Forest digunakan dan dinilai dengan menggunakan AUC. Seperti yang direferensikan, perhitungan pengaturan Rough Set mengalahkan perhitungan yang berbeda seperti Jaringan Saraf Persepsi Terpilih, Pohon Keputusan, dan Regresi Linier. enam sistem pengujian khas dianalisis oleh Amin et al. untuk oversampling masalah perkiraan agitasi telekomunikasi. Hasil menunjukkan bahwa MTDF dan aturan-usia yang bergantung pada perhitungan turun-temurun mengalahkan perhitungan oversampling lain yang dilihat sebelumnya.

Model pembelajaran mesin adalah representasi matematis. Itu dapat direpresentasikan ke proses dunia nyata. Pada data latih, algoritma pembelajaran ini mencari/menemukan pola sehingga parameter masukan akan sesuai dengan target yang diberikan. Prediksi dapat dibuat dengan menggunakan output dari proses pelatihan yang menghasilkan model pembelajaran mesin. Pelatihan model berarti belajar (menentukan) nilai-nilai yang baik untuk semua contoh berlabel yaitu. memiliki nilai bobot yang baik dan bias dari contoh berlabel. Dalam metode pembelajaran terawasi, algoritma pembelajaran mesin akan membangun model dengan memeriksa banyak contoh. Ia mencoba menemukan model yang akan menghasilkan minimalisasi kerugian. Proses ini disebut sebagai minimisasi risiko empiris.

Logistic Regression

Pemeriksaan kekambuhan adalah jenis metode menampilkan prescient yang mengeksplorasi hubungan antara yang membutuhkan (target) dan variabel otonom (indikator). Prosedur ini digunakan untuk mengantisipasi, menunjukkan pengaturan waktu dan menemukan hubungan sebab akibat antara faktor-faktor tersebut. 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. 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.

Random Forest

Hutan acak adalah salah satu jenis algoritma pembelajaran terawasi yang digunakan untuk regresi serta klasifikasi. Tapi itu terutama untuk klasifikasi jenis masalah. Hutan terdiri dari pohon dan lebih banyak pohon berarti hutan itu lebih kuat. Demikian pula, di sini algoritma hutan acak membuat pohon keputusan pada sampel data dan kemudian mendapatkan prediksi dari masing-masingnya dan akhirnya solusi terbaik dipilih dengan pemungutan suara. Ini adalah salah satu metode ensemble yang lebih baik daripada pohon keputusan tunggal karena akan mengurangi over-fitting dengan hasil rata-rata.

A Data Mining Approach to Customer Segment Based on Customer Value by Yun Chen, Chuan Fu, Hanhong Zhu

Segmentasi pelanggan adalah masalah dasar untuk CRM yang efektif, dan metode serta indikator segmentasi akan berdampak pada hasil aplikasi. Banyak literatur telah meneliti penerapan teknologi data mining dalam segmentasi pelanggan, dan mencapai hasil yang efektif. Tetapi kebanyakan dari mereka hanya mengelompokkan pelanggan dengan teknologi penambangan data tunggal dari pandangan khusus, bukan dari kerangka kerja sistematis. Meskipun metode segmen sebelumnya dapat mengelompokkan basis pelanggan ke dalam kelompok yang berbeda, metode ini tidak dapat mengidentifikasi kemampuan pelanggan untuk menciptakan keuntungan secara langsung. Dalam makalah ini, kerangka segmentasi pelanggan berdasarkan data mining diajukan, dan metode segmentasi pelanggan baru disusun berdasarkan Nilai Seumur Hidup Pelanggan. Metode yang diperkenalkan di atas telah diterapkan pada kumpulan data dari Frequent Flyer Program (FFP) sebuah maskapai penerbangan di China.

Menurut Newell (2000) CRM adalah alat yang berguna dalam hal mengidentifikasi kelompok pelanggan yang tepat dan membantu memutuskan pelanggan mana yang akan dibuang. Clemons (2000) memperkirakan mungkin ada perbedaan sepuluh kali lipat antara pelanggan yang paling menguntungkan dan pelanggan rata-rata. Penelitian telah menunjukkan bahwa penerapan

CRM menghasilkan kinerja perusahaan yang lebih baik ketika manajer fokus pada memaksimalkan nilai pelanggan (Gupta, Sunil, Donald R. Lehmann, dan Jennifer A. Stuart, 2004). Semakin banyak pemahaman pelanggan, semakin banyak nilai yang diperoleh dari mereka. Umumnya, metode segmentasi pelanggan sebagian besar mencakup metode deskripsi pengalaman, metode statistik tradisional, dan metode non-statistik (Per Vagn Freytag, et al, 2001, Lei-da Chen et al., 2000). Metode nonstatistik, mis. teknologi data mining, terutama muncul metode dalam segmentasi pelanggan (Agnes Nairn, dan Paul Bottomley, 2003, 2003, Jon Kleinberg, et al., 2004). Literatur ini menggunakan teknologi data mining tunggal untuk menganalisis masalah bisnis tunggal, dan telah mendapatkan beberapa hasil yang baik. Tetapi aplikasi ini memiliki satu kekurangan yang jelas yaitu metode segmentasi tersebut mungkin dapat mengelompokkan pelanggan ke dalam kelompok-kelompok tertentu dengan atribut yang berbeda, yang merupakan indikator tidak langsung daripada langsung. Upaya penelitian kami ditujukan pada sintesis konsep nilai pelanggan dengan mensegmentasi pelanggan oleh CLV, indikator langsung yang mewakili nilai moneter dari

pelanggan. Makalah ini berusaha untuk secara formal memasukkan konsep nilai pelanggan ke dalam manajemen hubungan pelanggan dengan mengembangkan kerangka segmentasi pelanggan yang terintegrasi. Tidak hanya menyediakan cara untuk memandu teknologi data mining yang digunakan dalam segmentasi pelanggan, tetapi juga mewujudkan model segmen pelanggan baru yang disebut Metode Segmentasi Pelanggan Berdasarkan CLV (CSMBC). Perusahaan dapat mengelompokkan pelanggan secara efektif dan mengelola kelompok pelanggan yang berbeda secara terpisah berdasarkan nilai masing-masing kelompok.

Customer Segment

Sebagai praktisi dengan antusias mencari kelompok pelanggan yang menguntungkan dengan CLV tinggi, beberapa akademisi mulai mempertanyakan apakah segmen sebenarnya adalah entitas yang stabil dan apakah mereka benar-benar ada (Jon Kleinberg, et al., 2004). Segmentasi pelanggan berdasarkan data mining dapat mengatasi masalah ini karena cara-cara tersebut dapat mempelajari informasi baru yang masuk kemudian dan mendapatkan aturan baru. Ini memberikan dukungan penuh untuk proses manajemen dinamis pencapaian pelanggan, retensi dan win-back, dan membangun hubungan pemetaan antara atribut pelanggan,

atribut dispersif dan kontinu. Menetapkan setiap atribut pelanggan sebagai dimensi dan menetapkan setiap pelanggan sebagai partikel, seluruh pelanggan dalam suatu perusahaan dapat membentuk ruang multidimensi, yang telah didefinisikan sebagai ruang atribut pelanggan. Secara umum, data mining adalah proses pencarian informasi yang sebelumnya tidak diketahui tetapi bermakna, seperti pola dan tren pengambilan keputusan, dengan menyaring kumpulan data yang besar dan memanfaatkan kombinasi pengenalan pola, pembuatan model, dan teknik validasi. Dalam makalah ini, kami memilih pengelompokan dari berbagai metode penambangan data karena daya tarik visualnya dan kesederhanaan dalam mengelompokkan objek ke dalam kelompok.

Segmentasi pasar dengan pola konsumsi bisa cukup berwawasan luas untuk memahami basis pelanggan. Manajemen hubungan pelanggan yang berbeda strategi yang sesuai untuk berbagai kelompok pengguna, seperti yang ditunjukkan pada gambar 2. Dengan mengelompokkan rekening nasabah berdasarkan CLV, perusahaan dapat mengembangkan strategi yang efektif dan menguntungkan untuk mempertahankan dan meningkatkan pelanggan.

Jelas, cluster 1 terdiri dari CLV tinggi yang berada dalam fase stabil churn dengan loyalitas tinggi. Orang-orang ini harus puas dengan produk atau layanan dan berniat untuk mempertahankan hubungan jangka panjang dengan perusahaan. Mereka adalah sapi perah bagi perusahaan dan tidak perlu menyuntikkan biaya besar untuk mempertahankannya, tetapi hanya beberapa kontak biasa. Cluster 2 mengandung CLV yang berisiko tinggi, dan perusahaan harus memberikan banyak perhatian kepada mereka yang mungkin menjadi bintang masa depan. Mereka mungkin adalah orang-orang yang ragu-ragu atas produk atau layanan dan belum membangun loyalitas kepada perusahaan. Oleh karena itu, perusahaan perlu menempatkan lebih banyak upaya pada pelanggan untuk mengatur hubungan yang panjang dengan mereka. Pelanggan setia tetapi CLV rendah merupakan klaster 3. Mereka mungkin merasa puas dengan produk atau layanan tetapi kemampuan ekonomi mereka relatif rendah. Mereka adalah bagian dari basis pelanggan perusahaan dan biaya yang memadai diperlukan untuk menjaga mereka agar tidak tertarik oleh pesaing. Cluster 4 terdiri dari

pelanggan yang nilainya rendah dan kepercayaan mereka belum dibangun. Orang-orang ini mungkin tidak puas dengan produk atau layanan atau sensitif terhadap harga. Perusahaan tidak perlu mengalokasikan banyak sumber daya untuk mereka karena mereka dapat memberikan pengaruh negatif bagi perusahaan. Umumnya, analisis semacam ini cocok untuk memahami perilaku pelanggan dengan nilai CLV yang berbeda. Strategi yang efektif ini mengklasifikasikan pelanggan melalui CLV untuk merancang manajemen hubungan pelanggan yang berbeda untuk segmentasi pelanggan yang berbeda. Singkatnya, model CLV dapat bermanfaat bagi aktivitas manajemen hubungan pelanggan. Sangat berguna untuk memikirkan cara mempertahankan pelanggan CLV tinggi, meningkatkan pengguna ringan dan menengah, membangun loyalitas pelanggan, memahami motif pembelian untuk memenuhi/melebihi harapan, menggunakan strategi penjualan yang tepat untuk setiap kelompok penggunaan yang ditargetkan, memenangkan kembali pelanggan yang hilang, dan mempelajari alasannya. bukan pengguna tidak menanggapi proposisi nilai.

Tujuan utama pemasaran adalah mengidentifikasi

pelanggan atau segmen mereka dengan penciptaan nilai terbesar

potensi dan targetkan mereka dengan sukses dengan

strategi pemasaran yang sesuai untuk mengurangi risiko

pelanggan dengan nilai seumur hidup yang tinggi ini membelot ke

pesaing (Andrew Banasiewicz, 2004). Segmentasi

pelanggan adalah pekerjaan dasar penambangan data menurut

informasi segmentasi historis yang diketahui. Pelatihan

data yang digunakan untuk membangun mode perkiraan segmen dapat

data historis atau data eksogen yang diperoleh dari

pengalaman atau survei.

Untuk perusahaan, bagaimana menggunakan teknologi data mining,

dan bagaimana memilih indikator segmen yang tepat sangat

penting, yang hasilnya akan berdampak pada perusahaan

menerapkan strategi manajemen hubungan pelanggan

secara langsung. Untuk menjawab pertanyaan ini, makalah ini mengusulkan

kerangka segmentasi berdasarkan data mining dan

membangun metode segmentasi berdasarkan CLV. Oleh

mengelompokkan pelanggan dengan waktu diskrit dan

indikator segmen waktu terus menerus, perusahaan dapat membuat

tindakan manajemen hubungan yang berbeda untuk masing-masing

segmen.

Metode segmen dalam makalah ini telah diterapkan untuk

industri penerbangan sipil, dan juga dapat digunakan dalam

industri lain seperti telekomunikasi, keuangan

layanan dll.

Predicting customer profitability over time based on RFM time series by Daqing Chen

Memprediksi profitabilitas konsumen secara dinamis dari waktu ke waktu memainkan peran penting dalam bisnis yang berpusat pada pelanggan saat ini. Dalam makalah ini, kami mengadopsi pendekatan sistem dinamis untuk mengatasi masalah prediksi dinamis profitabilitas pelanggan. Berdasarkan catatan transaksi pelanggan, rangkaian waktu berbasis skor RFM dihasilkan menggunakan analisis klaster. Deret waktu ini digunakan untuk mengukur dan menggambarkan profitabilitas pelanggan. Selanjutnya, model jaringan saraf umpan maju multilayer dilatih untuk menangkap dinamika profitabilitas pelanggan yang berkembang. Satu set transaksi nyata dari pengecer online yang berbasis di Inggris digunakan dalam penelitian ini. Hasil eksperimen yang relevan telah menunjukkan kinerja yang baik dari pendekatan yang diusulkan.

Memprediksi profitabilitas konsumen secara dinamis dari waktu ke waktu memainkan peran penting dalam bisnis yang berpusat pada pelanggan saat ini. Secara tradisional, bisnis biasanya memeriksa profitabilitas pelanggan pada titik batas waktu tertentu untuk menargetkan beberapa pelanggan tertentu untuk tujuan pemasaran. Namun pada kenyataannya, kebiasaan dan preferensi pembelian pelanggan bersifat heterogen dan dapat berfluktuasi dari waktu ke waktu. Oleh karena itu, bisnis ingin mengetahui bagaimana, berdasarkan pembelian historis dan profitabilitas setiap pelanggan, profitabilitasnya cenderung berkembang seiring waktu. Jawabannya akan berdampak langsung pada strategi pemasaran bisnis dan alokasi sumber daya. Secara umum, membuat prediksi tersebut memerlukan pertimbangan beberapa aspek utama, antara lain:

 Tentukan metrik untuk mengukur profitabilitas konsumen.

 Mengingat catatan riwayat pembelian konsumen dan metrik profitabilitas yang ditentukan, buat model dinamis yang sesuai untuk menggambarkan dinamika profitabilitas pelanggan, dan selanjutnya untuk memprediksi profitabilitas konsumen. Banyak metrik telah diperkenalkan dalam konteks bisnis, termasuk model Kekinian, Frekuensi, dan Moneter (RFM), model Nilai Seumur Hidup Pelanggan (CLV), dan model Ekuitas Pelanggan (CE) (Hughes, 2012; Kumar dan Reinartz, 2006). Setiap metrik memberikan perspektif unik dari mana bisnis dapat mengevaluasi perilaku pelanggan dan profitabilitas pelanggan. Menariknya, beberapa hubungan eksplisit atau implisit ada di antara metrik yang berbeda. Pemilihan metrik tertentu biasanya ditentukan oleh bisnis yang bersangkutan, dan terkadang beberapa metrik dapat diterapkan. Masalah utama yang dihadapi prediksi seperti itu biasanya terkait dengan pemodelan: mempertimbangkan profitabilitas konsumen sebagai proses yang berubah-ubah waktu, model mana yang paling sesuai untuk secara akurat menangkap karakteristik dinamis yang mendasari proses yang berkembang? Berbagai teknik pemodelan tersedia, misalnya analisis deret waktu (Tsay, 2010), jaringan Bayesian dinamis (Murphy, 2002), dan Jaringan Syaraf Tiruan (JST) (McNelis, 2005).

Dalam makalah ini, kami membahas masalah prediksi dinamis profitabilitas pelanggan dengan:

membangun model sistem dinamis, dan jaringan saraf umpan maju multilayer

(MFNN) dilatih untuk menangkap dinamika profitabilitas pelanggan yang terus berkembang. Satu set

catatan transaksi dari pengecer online yang berbasis di Inggris diselidiki. Metrik RFM

dipilih sebagai ukuran profitabilitas pelanggan. Dalam periode waktu tertentu, pelanggan adalah

dipartisi ke dalam kelompok dengan nilai RFM. Oleh karena itu, setiap pelanggan diberikan keunikan

Skor RFM untuk mencerminkan secara kolektif perilaku pembelian pelanggan dan profitabilitas. Sebagai

demikian, rangkaian waktu nilai profitabilitas pelanggan dengan tag periode waktu tertentu adalah

dibentuk dan digunakan untuk melatih dan menguji model jaringan syaraf tiruan. Performa dari

model MFNN yang dibangun diperiksa lebih lanjut sebagai model otonom, yaitu bagaimana

berkembang sebagai sistem umpan balik dinamis yang diberikan profitabilitas pelanggan awal.

Sisa dari makalah ini disusun sebagai berikut. Bagian 2 menyediakan tinjauan literatur

pekerjaan yang relevan, dengan fokus pada pendekatan dan model yang digunakan untuk pelanggan dinamis

prediksi profitabilitas. Bagian 3 menguraikan metodologi penelitian ini. Sebuah rinci

pembahasan diberikan mengenai penggunaan k-means clustering analysis untuk segmentasi

nilai RFM berbasis pelanggan, dan kumpulan data target yang akan diperiksa disajikan. Di dalam

Bagian 4, model MFNN dilatih untuk memprediksi profitabilitas pelanggan. Resultannya

model juga digunakan sebagai model otonom. Penampilan model-model ini adalah

dievaluasi dalam hal akurasi prediksi dan matriks kebingungan. Bagian 5 merangkum

temuan penting dari penelitian ini berdasarkan hasil eksperimen. Akhirnya, menyimpulkan

komentar dan garis besar pekerjaan masa depan diberikan dalam Bagian 6.

Literature

Memprediksi profitabilitas pelanggan secara akurat dari waktu ke waktu berdasarkan pembelian historis catatan telah terbukti menantang dan 'belum sepenuhnya berhasil' (Rusta et al.,

2011). Hal ini terutama disebabkan oleh fakta bahwa selalu ada tingkat ketidakpastian tertentu

terkait dengan pembelian pelanggan (Malthouse dan Blattberg, 2005), dan pelanggan yang paling menguntungkan selama satu periode waktu mungkin menjadi yang paling tidak menguntungkan selama periode waktu lain. Selain itu, dalam pengaturan non-kontraktual tipikal di B2C atau B2B

lingkungan, sangat umum bahwa satu-satunya informasi yang dimiliki perusahaan tentang

pelanggan adalah pembelian mereka (dan dalam beberapa kasus, alamat pengiriman dan penagihan mereka, mis.

untuk pengecer online). Perusahaan hampir tidak tahu apakah keadaan pelanggan telah berubah

waktu, dalam hal, misalnya, peran pekerjaan atau gaji mereka, dll.

demografi membuatnya semakin sulit untuk memprediksi profitabilitas pelanggan secara akurat

dan secara dinamis. Oleh karena itu, banyak teori probabilitas dan proses stokastik berbasis

model telah diusulkan untuk mengatasi masalah prediksi dinamis masa depan

profitabilitas pelanggan. Beberapa pekerjaan sebelumnya yang berkaitan dengan aspek ini dibahas di bawah ini.

Model Rantai Markov (MCM) telah digunakan secara luas dalam memperkirakan CLV dan

memodelkan hubungan pelanggan secara umum (Pfeifer dan Carraway, 2000; Bozzetto et al.,

2005; So dan Thomas, 2011). Dalam penelitian Haa et al. (2002), setiap pelanggan

profitabilitas atau 'status' pada waktu tertentu t diwakili oleh MCM, dan yang terkait

matriks transisi negara digunakan untuk menggambarkan bagaimana segmen pelanggan berbasis RFM memiliki

berkembang dari waktu ke waktu dalam basis pelanggan tertentu. Pengetahuan ini selanjutnya terkait dengan dinamika

profitabilitas pelanggan dan strategi pemasaran yang tepat.

Lebih umum, satu set model probabilistik telah dibahas untuk mengukur masa depan

profitabilitas pelanggan dalam hal laba bersih dan nilai sekarang bersih dari laba (Rusta et al., 2011). Model termasuk model probabilitas bersama dari laba kotor, pembelian pelanggan

insiden, dan kontak pasar. Model-model ini, meskipun rumit, luar biasa

mengungguli model naif. Berdasarkan hasil simulasi, Rusta et al. (2011) juga berpendapat

bahwa 'memprediksi profitabilitas pelanggan di masa depan tidak sia-sia; itu membutuhkan canggih

pendekatan pemodelan'.

Keuntungan utama dari model yang diperiksa di atas adalah bahwa mereka mewakili kotak putih

pendekatan untuk masalah dan dapat digunakan untuk membangun hubungan eksplisit antara

berbagai variabel yang terlibat yang didasarkan pada probabilitas dan proses stokastik

teori. Oleh karena itu, model-model ini mudah dipahami dan diinterpretasikan. Di samping itu,

Namun, model ini bergantung pada sampel independen dan representatif yang cukup dari a

basis data pelanggan untuk menghasilkan perkiraan yang andal dari berbagai distribusi probabilitas sebagai

diperlukan, dan biasanya membutuhkan komputasi yang intensif. Selain itu, model ini hanya dapat

beradaptasi dengan proses stokastik yang berbeda sampai batas tertentu.

Machine learning based classification and segmentation techniques for CRM: a customer analytics by Narenda Singh

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).

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.

Revolusi digital dan peningkatan jumlah data yang dihasilkan oleh perusahaan/organisasi dalam beberapa dekade terakhir telah menyebabkan minat yang besar di bidang pembelajaran mesin (ML) dan pembelajaran mendalam. Organisasi dan perusahaan teknologi menggunakan analitik prediktif berbasis ML 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. Teknik ML 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) (Cioca et al., 2013; Rahman dan Khan, 2017). Dalam skenario saat ini, proses bisnis menjadi semakin berorientasi pada pelanggan dan ditempatkan sebagai prioritas utama manajemen. Karena kemajuan teknologi dalam e-commerce, M-commerce, pemasaran virtual dan pemasaran digital, setiap produk hanya berjarak satu klik dari pelanggan. Hal ini menyebabkan persaingan yang sangat ketat, yang diperlukan untuk memastikan bahwa konsumen menerima standar kualitas setinggi mungkin untuk mempertahankannya (Singh dan Agrawal, 2019a; Adebiyi et al., 2016) dan mengurangi tingkat pengadukan. 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. Juga sangat penting untuk mengamati faktor-faktor kunci terpenting yang mempengaruhi keputusan pembelian pelanggan untuk membeli produk dan layanan apa pun (Singh et al., 2018; Yadav et al., 2018). ML 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. Segmentasi pelanggan berguna dalam memahami sub populasi demografis dan psikografis apa yang ada di dalam pelanggan Anda dalam kasus bisnis (Sgaier et al. , 2017) dan memanfaatkan informasi ini untuk meningkatkan laba, citra, nilai, dan manajemen inventaris. 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. 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.

Tujuan dasar dari CRM adalah untuk menyarankan bahwa pembeli harus mengidentifikasi pelanggan mereka yang paling menguntungkan dan harus fokus untuk membangun hubungan dengan mereka (Malmi et al., 2004). Penerapan teknik data mining dalam CRM dilakukan dengan menggunakan klasifikasi, clustering, dan prediksi berdasarkan pengetahuan pelanggan (Ngai et al., 2009). Kerangka kerja penambangan data CRM yang efisien digunakan untuk prediksi perilaku pelanggan untuk memperkaya proses pengambilan keputusan untuk retensi pelanggan (Bahari dan Elayidom, 2015). Teknologi selalu membantu untuk memahami persepsi pelanggan, kebutuhan dan harapan konsumen misalnya IoT di restoran (Sudhagar, 2019), Blockchain dalam hubungan konsumen (Boukis, 2019), ML dalam harga mentah (Mitra dan Banga, 2019), demografi pelanggan (Al-Zuabi et al., 2019), dll. Faktor kunci yang mempengaruhi CRM diperlukan untuk mengidentifikasi di bidang apa pun (Singh dan Gupta, 2020). Informasi tentang pelanggan tetap menjadi poin strategis utama dalam aplikasi yang berbeda. Ini membawa kebutuhan untuk mempertimbangkan alat canggih yang tersedia dari teknologi Big Data, yang telah berhasil digunakan di bidang lain seperti Bioinformatika, perawatan kesehatan, keuangan, atau perhotelan (George et al., 2014; Talón-Ballestero et al., 2018). IoT dan data besar menyediakan data yang sangat banyak untuk diproses (Ahmed et al., 2017). Konsep ML dengan metode ekonometrik dan berbasis teori diperlukan untuk mengolah data (Wedel dan Kannan, 2016). ML adalah teknik analisis data terkini yang memiliki kemampuan untuk belajar tanpa diprogram secara eksplisit dan tanpa campur tangan manusia (Alpaydin, 2014; Marsland, 2015). Sebuah studi kasus tentang penjualan B2B direpresentasikan untuk mengidentifikasi pelanggan baru dengan menggunakan pembelajaran K-means dan PU dengan adaptor hutan acak (Norlin dan Paulsrud, 2017). Klasifikasi pelanggan dalam kategori yang berbeda digunakan dalam proses pengambilan keputusan perusahaan atau organisasi. Klasifikasi Pelanggan menjadi 'setia' atau 'tidak setia' dibahas dalam Singh dan Agrawal (2018a, 2018b) dan memanfaatkan masalah alokasi saluran di jaringan nirkabel. Faktanya, untuk bertahan di pasar, operator telekomunikasi lebih mengabdikan diri untuk mempertahankan pelanggan yang berharga daripada mendapatkan pelanggan baru (Singh dan Agrawal, 2019b). KNN, SVM, Naïve Bayes dll, teknik pembelajaran mesin digunakan untuk klasifikasi pelanggan (Kotsiantis et al., 2007). Metode ML digunakan untuk memecahkan masalah yang menantang seperti prediksi customer churning di industri telekomunikasi (Vafeiadis et al., 2015; Ahmad et al., 2019). ML dapat mendukung dalam bantuan pelanggan untuk meningkatkan penjualan toko fashion (Fares et al., 2019), karena brand passion menjadi bagian penting dari perilaku konsumen untuk loyalitas, kepuasan dan kepercayaan (D'lima, 2018). Dengan memanfaatkan teknik ML dengan database CRM, adalah mungkin untuk menemukan pelanggan yang berharga. Teknik ML digunakan untuk mengevaluasi analisis segmentasi pelanggan dinamis untuk pengguna seluler (Dullaghan dan Rozaki, 2017). Teknik ML adalah alat untuk menganalisis perilaku pelanggan atau pengguna dalam bisnis apa pun yang ingin tetap signifikan dan mendapatkan usia yang kompetitif. Dalam tulisan ini, penulis mengusulkan segmentasi dan klasifikasi pelanggan yang mengunjungi mall untuk melakukan pembelian. Analisis demografis dapat digunakan oleh organisasi untuk menargetkan pelanggan tertentu dan meningkatkan keuntungan mereka secara keseluruhan. Untuk memenuhi tujuan tersebut, penulis bermaksud mengikuti kontribusi utama dalam makalah ini:

 segmentasi pelanggan pada properti demografis untuk menganalisis data pelanggan

 klasifikasi pelanggan melalui teknik ML yang berbeda untuk melakukan keputusan

membuat strategi dalam CRM.

A Proficient CRM-Machine Learning Framework for The Prediction of Customer Behaviour Random forest by Nithilia B

Dalam CRM (manajemen hubungan pelanggan), perangkat lunak CRM adalah kategori perangkat lunak yang mencakup serangkaian luas aplikasi yang dirancang untuk membantu bisnis mengelola banyak proses bisnis berikut: data pelanggan. interaksi pelanggan. mengakses informasi bisnis. Contoh sistem CRM mencakup platform yang dibangun untuk mengelola pemasaran, penjualan, layanan pelanggan, dan dukungan, semuanya terhubung untuk membantu perusahaan bekerja lebih efektif. Dengan sistem CRM, bisnis dapat menganalisis interaksi pelanggan dan meningkatkan hubungan pelanggan mereka. Model prediksi berbasis informasi yang menggunakan teknik pembelajaran mesin telah mendapatkan popularitas besar selama beberapa dekade terakhir. Model tersebut telah diterapkan di sejumlah domain seperti diagnosis medis, prediksi kejahatan, rating film, dll. Mirip dengan tren di industri telekomunikasi di mana model prediksi telah diterapkan untuk memprediksi pelanggan yang tidak puas yang cenderung mengubah penyedia layanan. Karena biaya keuangan yang sangat besar dari churn pelanggan di telekomunikasi, perusahaan dari seluruh dunia telah menganalisis berbagai faktor (seperti biaya panggilan, kualitas panggilan, waktu respons layanan pelanggan, dll.) menggunakan beberapa teknik pembelajaran mesin. Karya ini mengusulkan berbagai teknik ML untuk prediksi churn pelanggan.

Hutan acak adalah algoritma pembelajaran terawasi yang digunakan baik untuk klasifikasi maupun regresi. 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. 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. 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. 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. Proses prediksi menggunakan hutan acak sangat memakan waktu dibandingkan dengan algoritma lain.

Ada sejumlah besar penelitian terkait dengan prediksi churn yang mengatasi keterbatasan kumpulan data mereka dan mengusulkan berbagai algoritme dan teknik untuk mengidentifikasi pelanggan churn 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 Pohon Keputusan dan K-Nearest Neighbor.

Studi kami berfokus pada melakukan perbandingan komprehensif antara Pohon Keputusan dan algoritma klasifikasi K-Nearest Neighbor. Kedua algoritma memiliki data input yang sama dengan rasio pelatihan dan pengujian yang sama. Data input untuk kedua algoritma diproses melalui filter untuk mengurangi noise dan menghapus data yang tidak diinginkan. Setelah itu, data yang telah dibersihkan dibagi menjadi set pelatihan dan pengujian. Set pelatihan kemudian dimodelkan menggunakan algoritma untuk memberikan output yang diinginkan. Subbagian berikut akan memberikan rincian lebih lanjut tentang proses ini.