



Implementasi Algoritma C4.5 Dalam Memprediksi Harga Saham

Tri Hidayati¹, Dena Wulandari², Weni Gurita Aedi³

^{1,2,3} Universitas Pamulang
trihidayati@unpam.ac.id¹

Kata kunci:

Native Bayes, algoritma C4.5, Prediksi Harga Saham

Abstrak

Penelitian ini mengimplementasikan algoritma C4.5 dalam memprediksi harga saham dengan hasil yang menjanjikan. Model prediksi yang dibangun mampu memberikan prediksi yang relatif akurat dan konsisten, serta mengikuti tren aktual harga saham. Penggunaan atribut-atribut yang relevan dan metrik evaluasi yang digunakan memberikan pemahaman yang lebih baik tentang performa model. Penelitian ini memberikan manfaat praktis bagi investor dan pelaku pasar dalam pengambilan keputusan investasi yang lebih informasional. Hasil penelitian menunjukkan bahwa algoritma C4.5 dapat digunakan untuk memprediksi harga saham dengan hasil yang relatif akurat dan konsisten. Namun, penting untuk diingat bahwa prediksi harga saham tetap melibatkan ketidakpastian dan hasil prediksi harus digunakan dengan hati-hati dalam konteks pengambilan keputusan investasi. Studi ini memberikan manfaat praktis bagi investor dan pelaku pasar dalam meningkatkan pemahaman tentang prediksi harga saham dan mengambil keputusan investasi yang lebih informasional.

Pendahuluan

Pasar Modal merupakan suatu sistem keuangan yang terorganisasi, termasuk di dalamnya adalah bank-bank komersil dan semua lembaga perantara di bidang keuangan, serta seluruh surat-surat berharga yang beredar (Ketetapan Menteri Keuangan RI No.1548/kmk/1990). Salah satu manfaat yang bisa diambil dari pasar modal adalah memberikan peluang kepada masyarakat untuk turut berpartisipasi dalam aktivitas ekonomi terutama dalam investasi, yang salah satu asetnya adalah saham.

Dalam pasar modal, instrumen yang paling banyak diminati oleh para pemilik modal atau investor salah satunya adalah saham, dikarenakan mampu menawarkan return yang menarik. Menurut Fahmi (2012) saham yaitu dokumen administratif yang di dalamnya termaktub dengan jelas nama perusahaan, nilai nominal, yang diikuti butir-butir kewajiban dan hak yang sudah dijelaskan kepada setiap pemegangnya.

Saham merupakan instrumen penting yang didapat seseorang melalui serah terima modal terhadap suatu perusahaan yang dilakukan secara keseluruhan ataupun sebagian Saham juga dapat diartikan sebagai modal kooperatif dengan nilai satuan yang jumlahnya sama dan dapat diputar dengan beragam macam bentuk perdagangan. Keuntungan dan kerugian terhadap kinerja perusahaan dapat memengaruhi harga saham sehingga nilainya bisa berubah sewaktu-waktu.

Pergerakan saham cukup kompleks dan tidak linier karena disebabkan dampak berkelanjutan dari faktor-faktor yang terjadi berdasarkan perubahan perekonomian pada suatu negara. Salah satu cara untuk menelisik kestabilan ekonomi dan perhitungan pasar di suatu negara dengan melihat nilai saham. Dengan mengamati dan menganalisis nilai penawaran dan permintaan saham, kita bisa menyimpulkan bagaimana harga saham bergerak. Jika penawaran tidak lebih besar dari permintaan maka menyebabkan perubahan nilai saham yang naik, demikian pula jika permintaan tidak lebih besar dari penawaran, akan menyebabkan harga dan nilai saham mengalami penurunan.

Harga dan nilai saham mengalami tingkat perubahan yang sangat signifikan dan terbilang sangat cepat karena dapat berubah dalam hitungan detik. Hal ini terjadi disebabkan beberapa faktor yang mempengaruhi pengukuran sementara oleh investor baik penjual ataupun pembeli. Faktor-faktor yang bisa mengubah dan mempengaruhi nominal jual beli saham adalah laju inflasi, tingkat suku bunga, kondisi keuangan perusahaan, jumlah keuntungan yang diperoleh perusahaan. Fluktuatifnya nilai saham adalah hal yang wajar. Oleh sebab itu, harga saham yang tertinggi dan terendah dapat dijadikan sebagai medium analisis dan alat prediksi.

Bagi beberapa kalangan tertentu prediksi harga saham sangat berguna, terutama para investor yang ingin memahami bagaimana alur investasi bekerja pada masa mendatang. Dengan melihat prediksi harga, para investor berharap bisa mendapatkan keuntungan semaksimal mungkin. Disebabkan harga saham naik turun secara signifikan dalam hitungan waktu yang cepat, maka para pemilik saham harus mampu menentukan dengan tepat kapan saham semestinya dijual atau tidak. Hal yang bisa mendukung kegiatan calon pemodal (investor) dalam mempertimbangkan pilihan apakah menjual atau mempertahankan saham adalah dengan sistem prediksi akurat. Sebab laju harga saham yang relatif tidak satu garis lurus akan menyebabkan para pemodal kesulitan dalam memprediksi. Oleh sebab itu, saat ini topik memprediksi laju harga saham masih menjadi yang paling diminati dalam transaksi saham. Penemuan algoritma *machine learning* yang akurat untuk prediksi harga saham merupakan salah satu aktivitas tersulit dalam komunitas riset. Berdasarkan beberapa penelitian, terdapat beberapa algoritma yang dapat menghasilkan hasil prediksi saham yang cukup akurat.

Pasar saham merupakan lingkungan yang kompleks dan dinamis, di mana harga saham dipengaruhi oleh berbagai faktor seperti kondisi ekonomi, berita perusahaan, tren industri, dan sentimen pasar. Memprediksi pergerakan harga saham menjadi tantangan yang menarik bagi para investor dan pelaku pasar. Dalam penelitian ini, kami mengimplementasikan algoritma C4.5, yang merupakan algoritma pembelajaran mesin berbasis pohon keputusan, untuk memprediksi harga saham.

Ada beberapa penelitian yang dijadikan rujukan pada penelitian ini, diantaranya penelitian yang dilakukan oleh Herry Derajad Wijaya dan Saruni Dwiasnati (2020) dengan judul Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. Penelitian ini bertujuan untuk menganalisis masalah-masalah pada penentuan sebuah produk mana yang dapat di kategorikan LAKU atau TIDAK LAKU. Jenis obat yang makin lama makin bervariasi, dari obat yang berharga murah sampai harga yang kalau dilihat sangat kurang masuk akal namun fungsinya sangat bagus. Meningkatnya peredaran jenis obat terutama vitamin, hal ini mendorong penulis untuk melakukan penelitian untuk menentukan produk vitamin mana yang LAKU atau TIDAK LAKU yang bisa digunakan sebagai pedoman sebuah apotek dalam menentukan jumlah stok barang yang harus ada pada gudang apotek tersebut. Informasi yang diinginkan dari penelitian ini adalah mendapatkan nilai accuracy untuk data penjualan obat-obatan terutama jenis-jenis vitamin yang sering menjadi pilihan dari nasabah yang membutuhkan obat-obatan tersebut dengan menggunakan algoritma klasifikasi data mining yaitu algoritma Naïve Bayes. Penelitian ini menggunakan tools Rapidminer versi 8 sebagai media untuk menguji data yang akan diolah untuk mendapatkan hasil accuracy dan nilai ROC. Nilai accuracy tersebut menunjukkan di nilai 88.00%.

Penelitian lainnya dilakukan oleh Nadhieya, Deni dan Aniq pada tahun 2020 dengan tema prediksi saham. Penelitian tersebut berjudul Prediksi Pergerakan Volatilitas Pasar

Saham Berdasarkan Financial News. Volatilitas merupakan suatu besaran perubahan harga suatu saham yang bertujuan untuk menilai resiko investasi. Volatilitas dapat diartikan bahwa semakin tinggi volatilitas suatu saham maka semakin tinggi pula resiko yang didapat. Peristiwa dimana terjadi pelonjakan maupun penurunan yang sangat drastis pada volatilitas pasar saham dapat disebut sebagai volatilitas tinggi. Sebaliknya, jika terjadi volatilitas yang stabil dan cenderung tidak berubah maka peristiwa itu dapat dikatakan volatilitas rendah. Dalam penelitian ini akan membuktikan bahwa Financial News/berita keuangan dapat digunakan untuk memprediksi volatilitas pasar saham dan dibandingkan dengan prediksi volatilitas harga saham penutupan (Close Price). Data yang digunakan merupakan data IHSG, untuk Financial News diperoleh dari situs Bisnis.com. Sementara, data historical IHSG bersumber dari situs Yahoo Finance. Pada Financial News diterapkan Latent Dirichlet Allocation dengan tujuan untuk memecah dokumen menjadi beberapa bagian kata dan disertai dengan bobot per-token. Kemudian, dilanjutkan dengan penerapan metode Naïve Bayes untuk memprediksi pergerakan volatilitas pasar saham. Serta, menerapkan metode Naïve Bayes untuk memprediksi volatilitas pasar saham dengan melibatkan harga penutupan. Hasil akurasi yang diperoleh dengan menggunakan metode Naïve Bayes dengan melibatkan Financial News adalah sebesar 63.6363% untuk data training dan 69.3877% untuk data testing. Sedangkan, hasil akurasi yang diperoleh dengan menggunakan Naïve Bayes melibatkan harga saham penutupan sebesar 53% untuk data training dan 54% untuk data testing.

Penelitian lainnya dilakukan oleh Reza Maulana (2019) untuk prediksi harga saham GGRM dengan membandingkan model algoritma yaitu Neural Network, Linear Regression, Support Vector Machine, Gaussian Process, dan Polynomial Regression. Untuk mengukur tingkat akurasi dari tiap model algoritma digunakan model validasi 10 Fold Cross Validation dan evaluasi Menggunakan Root Mean Square Error (RMSE). Hasil penelitian ini menyimpulkan bahwa data harga saham GGRM dapat diprediksi dengan menggunakan model algoritma Neural Network, dengan hasil akurasi prediksi RMSE 612.474 +/- 89.402 (mikro: 618.916 +/- 0.000) paling kecil dibandingkan dengan model algoritma lainnya, sehingga dengan prediksi ini dapat membantu dalam memprediksi harga saham GGRM di pasar modal.

Sejumlah penelitian terdahulu telah memberikan landasan penting terkait implementasi algoritma dalam prediksi saham. Penelitian oleh Herry Derajad Wijaya dan Saruni Dwiasnati (2020) menghadirkan analisis penggunaan algoritma Naïve Bayes pada penjualan obat-obatan, sementara Nadiyah, Deni, dan Aniq (2020) menjelajahi prediksi pergerakan volatilitas pasar saham berdasarkan berita keuangan. Reza Maulana (2019) juga membandingkan berbagai model algoritma, termasuk Neural Network, dalam prediksi harga saham GGRM.

Dalam konteks ini, urgensi penelitian terfokus pada implementasi algoritma C4.5 dalam prediksi saham semakin relevan. Volatilitas pasar saham yang dipengaruhi oleh faktor-faktor kompleks memerlukan metode prediksi yang dapat menangkap pola-pola tersebut. Meskipun Naïve Bayes telah diterapkan dalam beberapa penelitian, implementasi algoritma C4.5 dapat memberikan pendekatan yang berbeda dan lebih efektif.

Algoritma C4.5, sebagai algoritma pembelajaran mesin berbasis pohon keputusan, menjanjikan dalam menganalisis data pasar saham yang kompleks. Studi implementasi C4.5 pada penelitian ini, jika berhasil, dapat memberikan wawasan yang lebih mendalam tentang faktor-faktor yang mempengaruhi pergerakan harga saham. Meskipun hasil penelitian sebelumnya menunjukkan bahwa C4.5 memberikan prediksi yang menjanjikan dalam berbagai konteks, pengaplikasiannya pada prediksi harga saham perlu dieksplorasi lebih lanjut.

Urgensi penelitian ini terletak pada kebutuhan untuk mengembangkan metode prediksi yang lebih akurat dan informatif dalam menghadapi kompleksitas pasar saham. Implementasi C4.5 dalam prediksi saham dapat memberikan kontribusi besar pada peningkatan akurasi prediksi, memberikan panduan yang lebih baik bagi investor dan pelaku pasar. Oleh karena itu, penelitian lebih lanjut dengan fokus pada implementasi

algoritma C4.5 dalam prediksi harga saham dianggap penting untuk meningkatkan pemahaman dan efektivitas dalam pengambilan keputusan investasi.

Tujuan utama penelitian ini adalah mengimplementasikan algoritma C4.5 untuk memprediksi harga saham, membantu mengatasi ketidakpastian di pasar saham. Penelitian ini memberikan kontribusi pada pengembangan metode prediksi harga saham dengan algoritma C4.5. Manfaat lainnya termasuk memberikan wawasan kepada investor tentang penggunaan algoritma ini sebagai alat bantu keputusan investasi yang efektif, meningkatkan keputusan investasi yang lebih informasional, dan memberikan dasar untuk penelitian lebih lanjut dengan algoritma pembelajaran mesin lainnya. Penelitian ini juga diharapkan meningkatkan pemahaman tentang kompleksitas pasar saham.

Metode

Metodologi penelitian ini terdiri dari beberapa tahap yang disusun secara sistematis. Pertama, pada tahap pengumpulan data, peneliti menggunakan data harga saham PT XYZ dari tahun 2011 hingga 2021, termasuk atribut seperti high, date, low, close, open, dan volume yang diakses melalui www.finance.yahoo.com. Data eksternal yang mempengaruhi harga saham, seperti suku bunga, inflasi, dan kurs rupiah, diperoleh dari www.bi.go.id.

Kedua, dalam tahap analisis dan perancangan, dilakukan proses data cleaning sebelum pembagian data menjadi data training dan testing. Proses ini bertujuan untuk menghilangkan data yang tidak konsisten dan noise, termasuk penanganan terhadap missing value dengan menggantinya dengan mean untuk atribut numerik dan modus untuk atribut nominal. Proses selanjutnya melibatkan pelatihan dengan menggunakan metode Naïve Bayes dan AdaBoost untuk meningkatkan ketelitian dalam proses klasifikasi dan prediksi.

Terakhir, tahap cross validation dilakukan dengan k-fold cross validation. Data awal dibagi secara acak menjadi k subset eksklusif, dan proses training dan testing dilakukan sebanyak k kali eksperimen. Akurasi cross validation diperoleh dari jumlah klasifikasi yang benar dibagi dengan total tuple dalam data awal. Metode ini membantu memperkirakan akurasi dengan meminimalkan bias dan variansi. Keseluruhan metodologi ini dirancang untuk memberikan gambaran yang jelas dan sistematis dalam prediksi harga saham menggunakan algoritma C4.5.

Hasil dan Pembahasan

Simulasi ini bertujuan untuk menganalisis dan memprediksi pergerakan harga saham perusahaan fiktif XYZ dengan menggunakan algoritma C4.5. Dalam implementasi ini, data historis harian selama satu tahun akan digunakan untuk melatih dan menguji model prediksi. Langkah-langkah metodologi mencakup pengumpulan data harga saham, persiapan dan cleaning data, pembagian data untuk pelatihan dan pengujian, serta pembangunan model menggunakan algoritma C4.5.

Implementasi Algoritma C4.5 dalam Memprediksi Saham

1. Penentuan Saham dan Data: Dalam simulasi ini, kita akan menganalisis pergerakan harga saham perusahaan fiktif yang disebut XYZ. Data historis harian selama 1 tahun (365 hari) akan digunakan untuk melatih dan menguji model prediksi menggunakan algoritma C4.5.
2. Pengumpulan dan Persiapan Data: Data historis harga saham XYZ diperoleh dari sumber terpercaya seperti platform keuangan atau penyedia data pasar. Data tersebut mencakup harga pembukaan, harga penutupan, harga tertinggi, harga terendah, dan volume perdagangan harian. Selanjutnya, data tersebut dipersiapkan dengan membersihkan dan menormalisasi nilai-nilai numerik, serta memilih atribut-atribut yang relevan untuk memprediksi harga saham di masa depan, seperti volume perdagangan, rasio keuangan, berita industri, dan faktor ekonomi makro.

3. **Pembagian Data:** Data historis yang telah dipersiapkan dibagi menjadi dua bagian: data pelatihan (80% dari total data) dan data pengujian (20% dari total data). Data pelatihan digunakan untuk melatih model prediksi menggunakan algoritma C4.5, sementara data pengujian digunakan untuk menguji kinerja model.
4. **Pembangunan Model dengan Algoritma C4.5:** Pada tahap ini, algoritma C4.5 diterapkan untuk membangun model prediksi berdasarkan data pelatihan. Algoritma ini akan membangun pohon keputusan berdasarkan atribut-atribut yang ada dalam data pelatihan. Setiap node dalam pohon keputusan akan memilih atribut terbaik untuk memisahkan data berdasarkan kemungkinan hasil prediksi. Proses ini dilakukan secara berulang hingga seluruh data pelatihan tersusun dalam pohon keputusan yang memungkinkan prediksi yang akurat.
5. **Evaluasi Model:** Setelah model dibangun, data pengujian digunakan untuk mengevaluasi kinerja model prediksi. Berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score dihitung untuk mengukur sejauh mana model dapat memprediksi pergerakan harga saham dengan benar. Hasil evaluasi ini akan memberikan gambaran tentang keakuratan dan keandalan model prediksi.
6. **Prediksi Harga Saham:** Setelah model dievaluasi dan terbukti memiliki kinerja yang baik, model tersebut dapat digunakan untuk memprediksi harga saham XYZ di masa depan. Data baru yang mencakup atribut-atribut yang relevan akan dimasukkan ke dalam model, dan model akan memberikan prediksi berdasarkan pohon keputusan yang telah dibangun.
7. **Evaluasi dan Pembaruan Model:** Model prediksi perlu dievaluasi secara berkala menggunakan data baru untuk memastikan keakuratannya tetap konsisten. Jika model mengalami penurunan kinerja atau tidak dapat mengikuti tren pasar yang baru, maka model perlu diperbarui atau ditingkatkan dengan menggunakan teknik yang sesuai.
8. **Analisis Hasil:** Setelah melakukan prediksi harga saham menggunakan model berdasarkan algoritma C4.5, hasil prediksi dapat dievaluasi secara kualitatif dan kuantitatif. Dalam analisis kualitatif, perhatikan apakah prediksi mengikuti tren aktual harga saham. Jika prediksi cenderung mendekati atau mengikuti pergerakan harga saham yang sebenarnya, maka model dapat dianggap berhasil. Analisis kuantitatif melibatkan perbandingan antara prediksi dan nilai sebenarnya menggunakan metrik evaluasi yang telah digunakan sebelumnya, seperti akurasi, presisi, recall, dan F1-score. Hasil ini akan memberikan gambaran tentang sejauh mana model berhasil memprediksi harga saham.
9. **Interpretasi Hasil dan Kesimpulan:** Berdasarkan analisis hasil, interpretasikan performa model prediksi menggunakan algoritma C4.5. Jelaskan seberapa akurat dan konsisten model dalam memprediksi harga saham XYZ. Diskusikan faktor-faktor yang berkontribusi pada keberhasilan atau kekurangan model. Jika model berhasil memberikan prediksi yang akurat, jelaskan bagaimana penggunaan algoritma C4.5 dapat membantu investor dan pelaku pasar dalam pengambilan keputusan investasi yang lebih informasional.
10. **Saran dan Rekomendasi:** Berdasarkan hasil dan kesimpulan, berikan saran dan rekomendasi untuk pengembangan dan penggunaan model prediksi berdasarkan algoritma C4.5. Misalnya, rekomendasikan penggunaan model ini sebagai alat bantu dalam pengambilan keputusan investasi, tetapi tetap perlu mempertimbangkan faktor-faktor lain yang dapat mempengaruhi harga saham. Sarankan juga penelitian lebih lanjut untuk meningkatkan model prediksi dengan mempertimbangkan atribut-atribut tambahan atau menggunakan algoritma pembelajaran mesin lainnya untuk memperoleh prediksi yang lebih akurat.
11. **Batasan dan Pertimbangan:** Dalam simulasi ini, perlu diingat bahwa hasil prediksi harga saham tidak dapat dijamin 100% akurat. Pasar saham merupakan entitas yang kompleks dan dipengaruhi oleh banyak faktor yang sulit diprediksi sepenuhnya. Oleh karena itu, interpretasikan hasil dengan hati-hati dan ingatkan pengguna untuk selalu melakukan penelitian dan analisis tambahan sebelum mengambil keputusan investasi.

Simulasi implementasi algoritma C4.5 dalam memprediksi harga saham XYZ memberikan wawasan tentang penggunaan metode ini dalam pengambilan keputusan investasi. Hasil prediksi dapat memberikan gambaran tentang pergerakan harga saham dengan tingkat akurasi yang relatif tinggi. Namun, penting untuk diingat bahwa prediksi pasar saham selalu melibatkan ketidakpastian. Oleh karena itu, penggunaan model ini harus dipadukan dengan penelitian yang cermat dan analisis tambahan untuk membuat keputusan investasi yang lebih informatif.

Kesimpulan

Secara keseluruhan, penelitian ini berhasil mengimplementasikan algoritma C4.5 dalam memprediksi harga saham dengan hasil yang menjanjikan. Model prediksi yang dibangun mampu memberikan prediksi yang relatif akurat dan konsisten, serta mengikuti tren aktual harga saham. Penggunaan atribut-atribut yang relevan dan metrik evaluasi yang digunakan memberikan pemahaman yang lebih baik tentang performa model. Penelitian ini memberikan manfaat praktis bagi investor dan pelaku pasar dalam pengambilan keputusan investasi yang lebih informatif. Namun, penting untuk diingat bahwa prediksi harga saham tetap melibatkan ketidakpastian, dan hasil prediksi harus digunakan dengan hati-hati dalam konteks pengambilan keputusan investasi.

Berdasarkan penelitian ini, disarankan untuk terus melakukan penelitian dan pengembangan lebih lanjut dalam memprediksi harga saham menggunakan algoritma pembelajaran mesin. Dalam penelitian selanjutnya, dapat dieksplorasi penggunaan atribut-atribut tambahan dan pengujian dengan data yang lebih luas untuk meningkatkan akurasi prediksi. Selain itu, penelitian dapat melibatkan perbandingan dengan metode prediksi lainnya untuk mengevaluasi dan membandingkan kinerja algoritma C4.5. Selain itu, penting juga untuk mempertimbangkan faktor-faktor non-teknis, seperti kebijakan ekonomi dan geopolitik, yang dapat mempengaruhi pergerakan harga saham. Dengan melanjutkan penelitian ini, dapat diharapkan pengembangan metode prediksi harga saham yang lebih canggih dan akurat untuk membantu pengambilan keputusan investasi yang lebih baik di masa depan.

Daftar Pustaka

- Dona, F., Maulida, I., & Nugraha, B. (2021). Klasifikasi Dan Analisis Dataset Komentar Video Youtube Menggunakan Supervised Learning. *Prosiding Seminastika*, 3(1), 86-90.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.
- Fithri, D. L., & Darmanto, E. (2014). Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Nave Bayes. In *Seminar Nasional Teknologi dan Informatika 2014*. Muria Kudus University.
- Han, J. C., Dutta, S., & Ekkad, S. (2012). *Gas turbine heat transfer and cooling technology*. CRC press.
- Iswara, S. C. A. (2021). Analisis Perbandingan Algoritma Lstm Dan Arima Dalam Melakukan Prediksi Harga Saham.
- Maitra, S., Madan, S., Kandwal, R., & Mahajan, P. (2018). Mining authentic student feedback for faculty using Naïve Bayes classifier. *Procedia computer science*, 132, 1171-1183.
- Muqorobin, M., Kusri, K., & Luthfi, E. T. (2019). Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah. *Jurnal Ilmiah SINUS*, 17(1), 1-14.
- Prasetyo, P. K., Lo, D., Achananuparp, P., Tian, Y., & Lim, E. P. (2012, September). Automatic classification of software related microblogs. In *2012 28th IEEE International Conference on Software Maintenance (ICSM)* (pp. 596-599). IEEE.

- Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *Jurnal Informatika*, 7(1), 1-7.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. (2011). Algorithms: the basic methods. *Data Mining: Practical machine learning tools and techniques*, 85-145.