
USING SVM (SUPPORT VECTOR MACHINE) METHOD TO PREDICT BEST SELLER BOOK IN AMAZON

Nengah Widya Utami^{1*}, Joshuael Juinor Soplantila²

¹ Sistem Informasi Akuntansi, STMIK Primakara

²Informatika, STMIK Primakara

Email: Joshuaelsoplantila@gmail.com

Abstrak: Amazon perusahaan teknologi multinasional Amerika yang berbasis di Seattle, Washington, yang berfokus pada e-commerce, komputasi awan, streaming digital, dan kecerdasan buatan Big Four bersama dengan Google, Apple, dan Facebook. Oleh karena itu perusahaan ini pasti memiliki data penjualan yang banyak dan beragam. Salah satunya adalah data penjualan buku. Dalam penelitian ini penulis tertarik melakukan analisis pada penjualan buku terlaris dengan mengimplementasikan data mining menggunakan algoritma SVM, dimana data tersebut diambil di Kaggle.com berupa data set..

Kata Kunci: Buku, Data Mining, SVM

Abstract: Amazon is an American multinational technology company based in Seattle, Washington, which focuses on e-commerce, cloud computing, digital streaming, and the Big Four artificial intelligence along with Google, Apple, and Facebook. Therefore this company must have many and varied sales data. One of them is book sales data. In this study, the author is interested in analyzing best-selling book sales by implementing data mining using the SVM algorithm, where the data is taken at Kaggle.com in the form of a data set.

Keywords: Books, Data Mining, SVM.

1. PENDAHULUAN

Pada masa ini sangat banyak buku –buku yang memiliki penjualan yang tinggi terutama pada perusahaan –perusahaan yang memiliki nama yang sudah besar seperti Amazon, akan tetapi memiliki kesulitan dalam menganalisis data buku –buku yang memiliki penjualan terbanyak atau terpopuler. Maka dari itu hampir setiap perusahaan memiliki orang untuk melakukan data mining dalam tiap penjualan pertahun atau perbulan yang akan terus bertambah.

Penggunaan alat bantu atau aplikasi sangat di perlukan dalam data mining salah satunya yaitu Orange, adalah aplikasi yang berfungsi dalam data mining yang bisa memproses data berjumlah puluhan atau ratusan data.

Amazon.com, Inc. adalah perusahaan teknologi multinasional Amerika yang berbasis di Seattle, Washington, yang berfokus pada e-commerce, komputasi awan, streaming digital, dan kecerdasan buatan. Perusahaan ini dianggap sebagai salah satu perusahaan teknologi Big Four

bersama dengan Google, Apple, dan Facebook.

Buku terpopuler, merupakan penjualan atau pembelian buku yang sering di beli atau di cari oleh pembeli dan dapat di temukan di hampir semua toko buku yang ada

Orange, adalah software open source (karena ini software yang kode sumbernya terbuka untuk diupdate, dipelajari, diubah, dan disebarluaskan). Untuk pengolahan Data Analytics atau Data Mining yang akan anda perlukan menganalisis suatu data.

2. TINJAUAN PUSTAKA

2.1. Amazon

Amazon.com, Inc. adalah perusahaan teknologi multinasional Amerika yang berbasis di Seattle, Washington, yang berfokus pada e-commerce, komputasi awan, streaming digital, dan kecerdasan buatan. Perusahaan ini dianggap sebagai salah satu perusahaan teknologi Big Four bersama dengan Google, Apple, dan Facebook.

2.2. Best Seller

Best Seller atau buku terlaris adalah buku atau media lain yang tercatat memiliki status penjualan terbaik, dengan daftar buku terlaris yang diterbitkan oleh surat kabar, majalah, dan rantai toko buku.

2.3. SVM

Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun non linear.

SVM merupakan salah satu metode klasifikasi dalam data mining. Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang pattern recognition. Sebagai salah satu metode pattern recognition, usia SVM terbilang masih relatif muda. Walaupun demikian, evaluasi kemampuannya dalam berbagai aplikasinya menempatkannya sebagai state of the art dalam pattern recognition, dan dewasa ini merupakan salah satu tema yang berkembang dengan pesat. SVM adalah metode learning machine yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. Tulisan ini membahas teori dasar SVM dan aplikasinya dalam bioinformatika, khususnya pada analisa ekspresi gen yang diperoleh dari analisa microarray.

Secara umum dalam SVM terdapat 2 proses klasifikasi yaitu:

1. Proses Training: dimana menggunakan training set yang sudah diketahui label – labelnya untuk membangun model atau fungsi tertentu.
2. Proses testing: pada proses ini yang akan diproses yaitu data – data yang digunakan untuk diuji keakuratannya.



Gambar 1. SVM Method Flowchart

3. METODE

Pada bagian ini penulis memiliki tahapan – tahapan yang harus dilakukan sehingga mendapat hasil yang diinginkan, yaitu:

1. Persiapan

Pada tahap pertama akan melakukan, research atau penidentifikasi masalah, data pendukung, tinjauan pustaka dan titik akhir dari penelitian ini atau hasil dari penelitian

2. Pengumpulan Data

Pada tahap kedua ini akan dilakukan data gathering atau mengumpulkan data yang diperlukan untuk melakukan implementasi algoritma SVM, dan pengujiannya. Penelitian data penjualan buku terlaris di amazon di ambil dari Kaggle.com yang berupa data set yang berasal dari pusdatin.kemendes.go.id - Pemangkasan data Pada data set yang tersedia berjumlah 850 data dan di pangkas atau di kurangi sehingga yang terpakai hanya 210 data.

3. Penerapan SVM

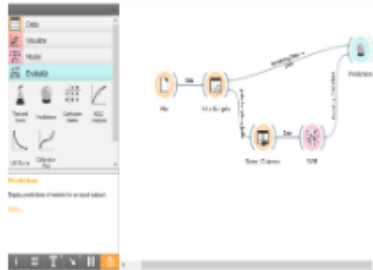
Setelah kedua langkah di atas telah dilakukan data yang sudah dipangkas akan diolah kedalam aplikasi yaitu Orange yang merupakan aplikasi data mining. Ketika telah diolah kedalam aplikasi mengikuti langkah –

langkah yang harus dilakukan maka akan muncul hasil dari banyaknya penjualan buku terbanyak pada Amazon

4. HASIL DAN PEMBAHASAN

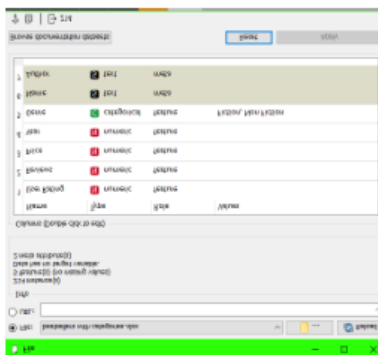
4.1. SVM (support vector system)

Data yang berjumlah 210 itu diproses di dalam Orange yang akan diproses seperti di bawah ini:

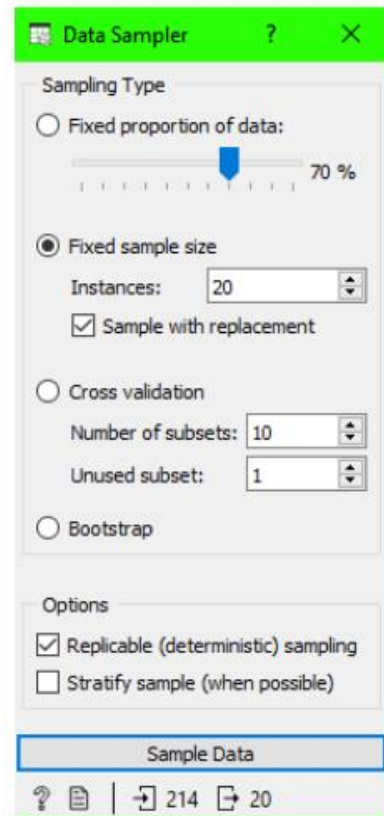


Gambar 2. Open file

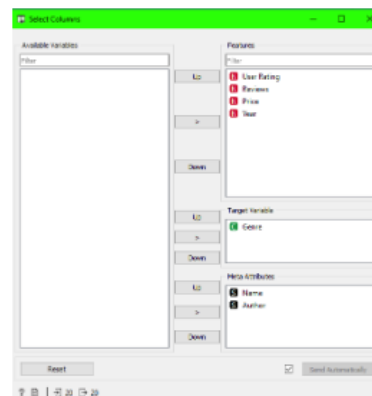
Setelah itu data yang akan diproses akan di masukan ke dalam poin –poin tertentu menggunakan tools yang tersedia di dalam Orange itu.



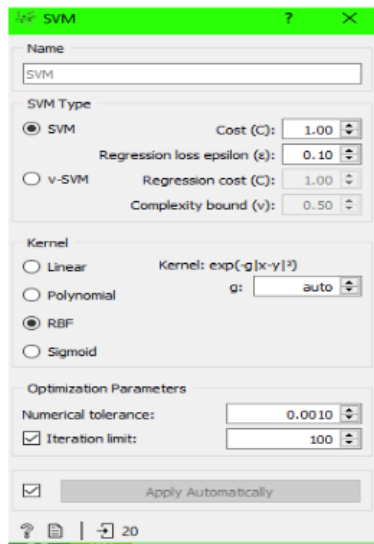
Gambar 3. File



Gambar 4. Data sampler



Gambar 5. Select column



Gambar 6. SVM

Genre	Title	Author	Book Rating	Reviews	Price	Year
Fiction	1770000-A...	Melanie King	4.5	250	35	2019
Fiction	1770000-B...	Andrew...	4.1	7670	36	2019
Fiction	1800000-A...	George Orwell	4.1	14500	9	2019
Fiction	1800000-B...	Melvin...	4.3	760	12	2019
Fiction	1770000-C...	George R. R...	4.1	1295	10	2019
Fiction	1770000-D...	George R. R...	4.1	7700	30	2019
Fiction	1770000-E...	Julia Todd	4.1	990	15	2019
Fiction	1770000-F...	James Comey	4.1	980	9	2019
Fiction	1770000-G...	Stephen King	4.4	7040	9	2019
Fiction	1770000-H...	Frank Richman	4.1	1000	9	2019
Fiction	1770000-I...	John Grisham	4.4	40	11	2019
Fiction	1770000-J...	James Joyce	4.1	480	32	2019
Fiction	1770000-K...	Michael...	4.1	580	9	2019
Fiction	1770000-L...	Michael...	4.1	110	4	2019
Fiction	1770000-M...	Blue Bear	4.4	210	9	2019
Fiction	1770000-N...	Michael...	4.4	240	9	2019
Fiction	1770000-O...	Colleen Hoover	4.1	200	9	2019
Fiction	1770000-P...	John Grisham	4.4	990	10	2019
Fiction	1770000-Q...	Colleen Hoover	4.1	2000	10	2019
Fiction	1770000-R...	Michael...	4.1	810	10	2019
Fiction	1770000-S...	Michael...	4.1	760	9	2019
Fiction	1770000-T...	Michael...	4.1	7100	10	2019
Fiction	1770000-U...	John Grisham	4.4	50	9	2019
Fiction	1770000-V...	Michael...	4.1	800	9	2019
Fiction	1770000-W...	Michael...	4.1	40	14	2019
Fiction	1770000-X...	Michael...	4.1	40	14	2019
Fiction	1770000-Y...	Michael...	4.1	240	9	2019
Fiction	1770000-Z...	Michael...	4.1	100	10	2019
Fiction	1770000-AA...	Michael...	4.1	100	10	2019

Gambar 7. Hasil

Sehingga pada akhir dari proses aplikasi ini akan muncul hasil seperti ini, dimana ada dua tipe buku yaitu Fiction dan non Fiction dimana tiap tipe memiliki buku yang masing – masing penjualan buku terpopuler.

5. KESIMPULAN DAN SARAN

Menurut pembahasan diatas, dapat disimpulkan bahwa hasil dari penelitian menunjukan 2 kriteria buku berbeda berdasarkan data yang pakai yaitu Fiction dan non Fiction, setiap kriteria memiliki buku yang memiliki tingkat penjualan tertinggi atau terpopuler, dan dapat dilihat ketika memilih salah satu kategori yang tersedia maka akan muncul presentase nilai tertinggi nya msing – masing

6. DAFTAR REFERENSI

- [1]. R, “Analisis Klasifikasi Menggunakan Metode SVM dari Data Penderita Penyakit Diabetes”, 2019.
- [2]. Samsudiney, “Penjelasan Sederhana tentang Apa Itu SVM”, 2019.