

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

Nengah Widya Utami, Joshuael Juinor Soplantila

Sistem Informasi Akuntansi, Teknik Informatika

STMIK Primakara

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 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 sumber-nya terbuka untuk diupdate, dipelajari, diubah, dan disebarluaskan). Untuk pengolahan Data Analytics atau Data Mining yang akan anda perlukan menganalisis suatu data.

TINJAUAN TEORITIS

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.

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.

SVM

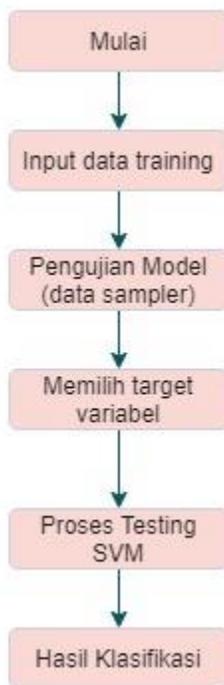
Apa itu SVM? Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas 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 kalsifikasi yaitu :

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



Gambar1. SVM Method Flowchart

METHOD

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

1. Persiapan

Pada tahap pertama akan melakukan, research atau penidentifikasian 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

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

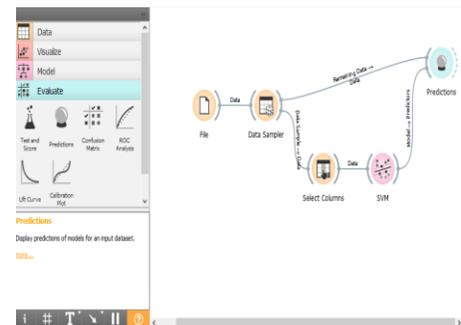
3. Penerapan SVM

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

TEMUAN DAN PEMBAHASAN

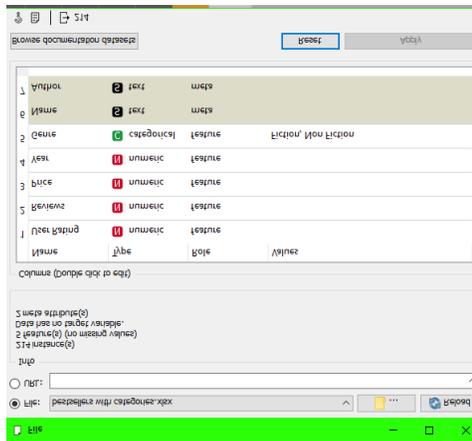
1. SVM (support vector system)

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

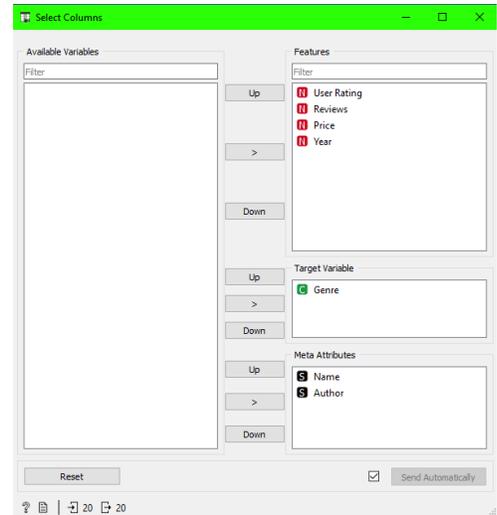


Gambar 2. Open file

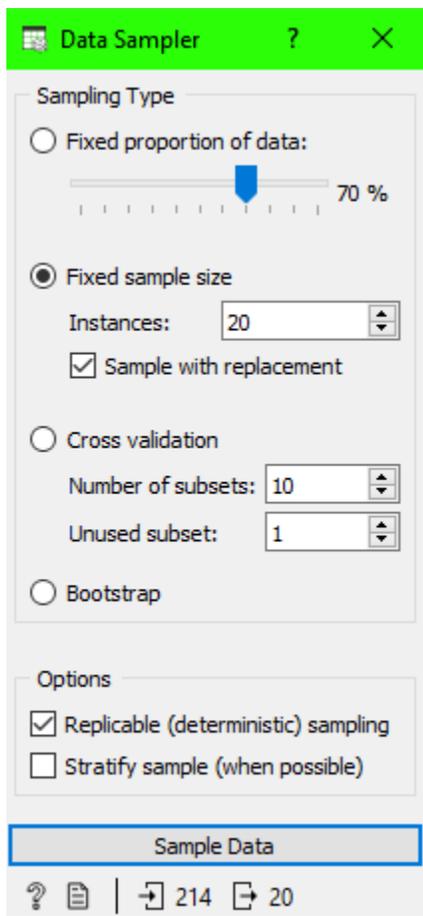
Setelah itu data yang akan diproses akan di masukan ke dalam poin – poin tertentu menggunakan tools yang tersedia di dalam *Orange* itu



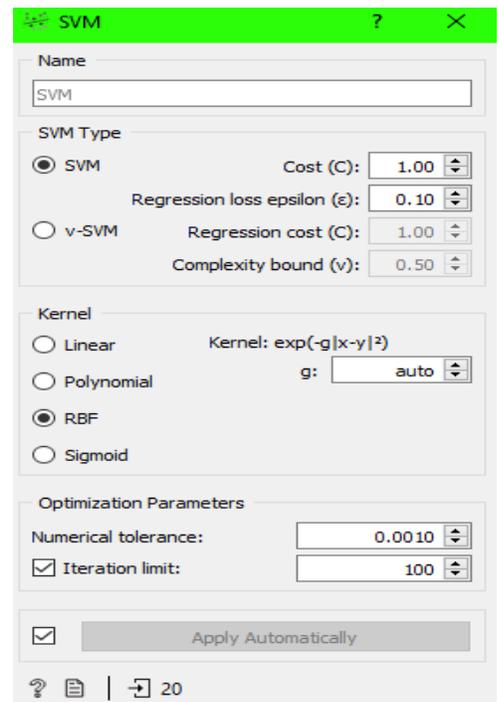
Gambar 3. File



Gambar 5. Select column



Gambar 4. Data sampler



Gambar 6. SVM

Rank	Genre	ISBN	Name	Author	User Rating	Reviews	Price	Year
1	Fiction	151215314	151215314	Stephen King	4.6	2152	12	2011
2	Non Fiction	110215314	110215314	Jordan B. ...	4.7	18079	15	2018
3	Fiction	12 Rules for Life	12 Rules for Life	Jordan B. ...	4.7	18079	15	2018
4	Non Fiction	1984 Dignat ...	1984 Dignat ...	George Orwell	4.7	21424	6	2017
5	Fiction	5000 Awesome...	5000 Awesome...	Natalie ...	4.3	7065	12	2019
6	Non Fiction	A Dance with ...	A Dance with ...	George R. R. ...	4.4	13043	11	2011
7	Fiction	A Game of ...	A Game of ...	George R. R. ...	4.7	19735	30	2014
8	Fiction	A Gentleman in ...	A Gentleman in ...	Amos Boules	4.7	19889	15	2017
9	Non Fiction	A Higher Loyalt...	A Higher Loyalt...	James Comey	4.7	5983	3	2018
10	Fiction	A Man Called ...	A Man Called ...	Fredrik Backman	4.6	23048	8	2016
11	Fiction	A Man Called ...	A Man Called ...	Fredrik Backman	4.6	23048	8	2017
12	Non Fiction	A Patriot's ...	A Patriot's ...	Larry Schweikart	4.6	460	2	2010
13	Non Fiction	A Stolen Life A ...	A Stolen Life A ...	Joyce Dugard	4.6	4149	32	2011
14	Non Fiction	A Mindlike T...	A Mindlike T...	Maddison...	4.5	13153	5	2018
15	Non Fiction	Adult Coloring ...	Adult Coloring ...	Adult Coloring ...	4.5	2313	4	2016
16	Non Fiction	Adult Coloring ...	Adult Coloring ...	Blue Star ...	4.6	2025	6	2015
17	Non Fiction	Adult Coloring ...	Adult Coloring ...	Blue Star ...	4.4	2031	6	2015
18	Non Fiction	Adult Coloring ...	Adult Coloring ...	Coloring Books	4.5	2426	8	2015
19	Fiction	Abuandar ...	Abuandar ...	Rui Chenour	4.8	9198	13	2016
20	Fiction	All the Light W...	All the Light W...	Anthony Doerr	4.6	16348	14	2015
21	Non Fiction	Allegiant	Allegiant	Veronica Roth	3.9	6310	13	2013
22	Non Fiction	American Spag...	American Spag...	Chris Kyle	4.6	15025	9	2015
23	Non Fiction	And the ...	And the ...	Khaled Hosseini	4.3	12159	13	2013
24	Non Fiction	Arguing with ...	Arguing with ...	Glenn Beck	4.6	798	5	2009
25	Non Fiction	Asymptotic Po...	Asymptotic Po...	Neil deGrasse	4.7	9374	9	2017
26	Non Fiction	Autobiography ...	Autobiography ...	Mark Twain	4.2	491	14	2010
27	Non Fiction	Baby Teach me...	Baby Teach me...	DK	4.6	5380	5	2015
28	Non Fiction	Balances (Anger...	Balances (Anger...	Angia Grace	4.8	1909	11	2015
29	Non Fiction	Bandages	Bandages	Bandages	4.8	1766	10	2013

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* di mana tiap tipe memiliki buku yang masing – masing penjualan buku terpopuler

KESIMPULAN

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 masing – masing.

Referensi

Islami, R. (2019, juni 7). *Analisis Klasifikasi Menggunakan Metode SVM dari Data Penderita Penyakit Diabetes*. Retrieved from medium.com:

<https://medium.com/@16611071/analisis-klasifikasi-menggunakan-metode-svm-dari-data-penderita-penyakit-diabetes-7329f73ef1da>

Samsudiney. (2019, juli 25). *Penjelasan Sederhana tentang Apa Itu SVM?* Retrieved from meduim.com:

<https://medium.com/@samsudiney/penjelasa-n-sederhana-tentang-apa-itu-svm-149fec72bd02#:~:text=Baiklah%20teman%2C%20kali%20ini%20saya,metode%20SVM%20dan%20sedikit%20ulasannya.&text=SVM%20digunakan%20untuk%20mencari,digunakan%20untuk%20pemisah%20antar%20kela>