

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut KBBI (Kamus Besar Bahasa Indonesia) ujian adalah sesuatu yang dipakai untuk menguji mutu sesuatu (kepandaian, kemampuan, hasil belajar dan sebagainya). Jadi hasil ujian tentunya mencerminkan tingkat kepandaian atau hasil belajar seseorang. Dalam penelitian ini penulis berharap menemukan faktor yang mempengaruhi hasil ujian dengan metode data mining. Dengan ditemukannya faktor tersebut maka dapat digunakan lembaga pendidikan untuk meningkatkan kualitas pengajarannya.

Support Vector Machine (SVM) merupakan salah satu teknik Data Mining untuk klasifikasi yang sangat terkenal dalam pattern classification dan machine learning karena kehandalannya dalam klasifikasi data multi dimensi (Kranjčić *et al.*, 2019). *Support Vector Machine (SVM)* adalah salah satu bagian dari Data Mining yang akhir-akhir ini mendapat perhatian dalam hal untuk pengklasifikasian data, selain Metode CART (*Classification And Regression Tree*), CHAID (*Chi Square Automatic Interaction Detection*) dan *k-NN (k-Nearest Neighbour)*. Salah satu kelebihan *SVM* dibandingkan metode lain terletak pada kemampuannya untuk menemukan hyperplane terbaik yang memisahkan dua buah class pada *Feature space* yang ditunjukkan oleh strategi *Structural Risk Minimization (SRM)* (Nurcahyo *et al.*, 2020). Metode pengklasifikasian menggunakan Support vector machine (*SVM*) merupakan suatu metode pattern recognition yang memetakan suatu data pada konsep tertentu yang telah di definisikan sebelumnya. Dalam hal ini *SVM* mengklasifikasikan ciri - ciri yang telah terekstraksi menjadi dua buah kelas sesuai dengan hyperplane terbaik (Putra, Purwandari and Desi, 2020).

Dalam studi *jaringan saraf tiruan (JST)*, *Neural Tangent Kernel (NTK)* adalah kernel yang menjelaskan evolusi jaringan saraf tiruan dalam selama pelatihan mereka dengan penurunan gradien. Hal ini memungkinkan JST untuk dipelajari menggunakan alat teoritis dari Metode Kernel. *NTK* diperkenalkan pada tahun 2018 oleh Arthur Jacot, Franck Gabriel dan Clément Hongler. (Jacot et al,2018).

Peneliti memilih memadukan metode *SVM* dan *NTK* karena seperti dijabarkan diatas *SVM* memiliki kelebihan dibanding metode lain dan *NTK* merupakan suatu teknik baru. *NTK* merupakan teknik yang baru ditemukan pada tahun 2018. Penulis sudah mencari di google scholar dan bantuan software Publish or Perish tetapi penggabungan *SVM* dan *NTK* untuk penelitian dalam bidang pendidikan juga tidak dapat penulis temukan. Pada masa sekarang ini metode *SVM* dan *NTK* dapat digabungkan untuk menambang data dengan bantuan software maupun program buatan sendiri. Sekarang ini banyak tersedia Software Data Mining yang apabila terdapat kekurangan dapat kita atasi dengan program buatan sendiri.

Dengan memadukan metode *SVM* dan *NTK* diharapkan dapat menghasilkan Prediksi hasil ujian siswa dengan tingkat keakuratan yang tinggi serta menyumbang pengetahuan baru bagi perpaduan metode data mining yang ada.

1.2 Masalah Penelitian

1.2.1 Identifikasi Masalah

Berikut ini akan dilakukan identifikasi masalah berdasarkan latar belakang yang telah diuraikan antara lain:

1. *SVM* lemah terhadap *Dataset* Besar. Sehingga penelitian dalam bidang pendidikan yang penulis temukan hampir semuanya menggunakan *Dataset* kecil untuk ditambang dengan metode *SVM*.
2. *NTK* merupakan metode yang baru saja ditemukan. Sehingga penulis hampir tidak bisa menemukan penelitian yang mengkombinasikan metode *SVM* dengan *NTK* sebagai trik kernelnya. Jika dibidang

pendidikan penulis belum menemukan satupun penelitian yang memprediksi hasil ujian dengan metode *SVM* dan *NTK*.

3. *Dataset Oulad* merupakan database pembelajaran online yang besar dan harus diolah terlebih dahulu baru dapat digunakan, untuk Indonesia pembelajaran online tidak begitu terkenal sehingga belum penulis temukan penelitian dengan *Dataset* pembelajaran Online.
4. Penelitian banyak menggunakan *Dataset* kecil karena di Indonesia pembelajaran online tidak begitu terkenal.

1.2.2 Rumusan Masalah

Detil permasalahan dalam penelitian ini dirumuskan sebagai berikut:

1. Mengkombinasikan metode *SVM* dan *Neural Tangent Kernel* untuk mendapatkan faktor yang mempengaruhi performa ujian dengan tingkat akurasi tinggi.
2. Mencari tingkatan pengaruh faktor yang mempengaruhi performa hasil ujian.
3. Menggunakan *Dataset* tersertifikasi pembelajaran online yang besar sebagai salah satu kontribusi mengikuti perkembangan terkini.

1.3 Tujuan Penelitian dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Beranjak dari latar belakang permasalahan dalam penelitian ini, secara detail tujuan penelitian ini dapat diuraikan sebagai berikut:

Untuk mengaplikasikan metode *SVM* dan *Neural Tangent Kernel* sebagai metode efektif dan efisien dalam proses memprediksi hasil ujian. Dari penelitian ini juga kita dapat mengetahui urutan faktor yang paling menentukan dari terbesar sampai terkecil dalam memprediksi keberhasilan hasil ujian siswa.

1.3.2 Manfaat Penelitian

Manfaat dari penelitian ini dapat dijabarkan sebagai berikut :

1. Untuk mengetahui faktor yang berpengaruh besar terhadap keberhasilan peserta didik menghadapi ujian. Dengan demikian badan pendidikan dapat mencegah siswa gagal menghadapi ujian yang berisiko siswa menjadi Drop Out.
2. Melihat lebih jauh hasil penelitian ini, diharapkan dapat menjadi model yang dapat diaplikasikan oleh badan pendidikan dalam meningkatkan kualitas pengajaran sehingga para siswa dapat memperoleh hasil ujian yang memuaskan
3. Penelitian ini secara teori ilmu pengetahuan diharapkan akan berkontribusi, menambah pengetahuan dan variasi penelitian dalam penerapan teknik data mining dan machine learning pada bidang pendidikan.
4. Selain itu hasil penelitian ini diharapkan dapat juga dijadikan sebagai perbandingan dan referensi bagi peneliti lain yang tertarik dengan teknik data mining pada bidang pendidikan, khususnya penerapan metode *SVM* dan *Neural Tangent Kernel*.
5. Penelitian yang telah ada banyak menggunakan *Dataset* kecil. Dimana seharusnya semakin banyak data training hasil yang didapatkan akan semakin bagus. Sehingga pada penelitian ini peneliti menggunakan *Dataset* yang besar.

1.4 Batasan Masalah

Mengingat luasnya ruang lingkup penelitian ini dan implementasi teknik-teknik data mining di area pendidikan, maka penelitian ini dibatasi pada:

1. Data penelitian ini, diperoleh dari database OULAD *Dataset* (Open University Learning Analytics *Dataset*).

2. Penelitian ini terbatas hanya untuk mengkaji metode *SVM* dan *Neural Tangent Kernel* dalam memprediksi performa hasil ujian siswa .
3. Penelitian ini tidak membandingkan hasil metode yang digunakan dalam penelitian ini yakni *SVM* dan *Neural Tangent Kernel* dengan metode - metode dalam prediksi hasil ujian yang lain.
4. Parameter yang digunakan dalam penelitian ini adalah
 - a. Demographic (Jenis Kelamin, Pendidikan Tertinggi, Umur).
 - b. Condition (Imd Band, Cacat atau tidak)
 - c. Performance (Kumpulan nilai yang didapat setiap siswa , dan Banyaknya usaha yang dilakukan).

1.5. Metodologi Penelitian

Berikut ini akan dijelaskan langkah - langkah metodologi penelitian yang digunakan antara lain:

1. Studi Literatur pada tahap ini dilakukan pendalaman topik yang diteliti dengan mempelajari cara kerja data mining khususnya *SVM* dan *Neural Tangent Kernel*.
2. Analisis pada tahap ini, dari hasil studi literatur yang sudah dilakukan identifikasi masalah dan tantangan yang harus diselesaikan dan menjelaskan solusi yang diusulkan untuk mengatasi masalah tersebut. Dalam hal ini kita akan menganalisis tingkat keakuratan prediksi .
3. Mempersiapkan data yang ingin ditambang. Dalam hal ini penulis akan mempersiapkan *Dataset Oulad* yang terdiri dari 22 kursus, 32.593 siswa dan hasil ujian mereka.
4. Mengaplikasikan metode *SVM* dan *Neural Tangent Kernel* untuk mengolah data dan mendapatkan hasil penelitian. Dalam hal ini kita akan menganalisis tingkat keakuratan prediksi serta waktu yang diperlukan oleh aplikasi *Rapid*

Miner untuk mengolah *Dataset* yang diberikan dengan metode yang disebutkan di point no 4.

5. Menguji hasil prediksi dengan *Cross Validation* dan *Confusion Matrix*.
6. Kesimpulan.

1.6. Sistematika Penulisan

Penulisan tesis ini terdiri dari 5 bab, dimana secara garis besar masing-masing bab membahas hal-hal berikut :

BAB I PENDAHULUAN

Bab ini merupakan pendahuluan yang berisi mengenai topik yang dikerjakan, menjelaskan latar belakang masalah, merumuskan masalah penelitian, tujuan dan manfaat yang diharapkan, ruang lingkup penelitian, metodologi penelitian serta sistematika penulisan.

BAB II KAJIAN LITERATUR

Bab ini berisi tinjauan pustaka dan landasan teori yang menunjang dalam pembuatan tesis. Dalam bab ini banyak penulis menggunakan referensi yang bersumber dari paper, buku dan lain-lain.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan tahapan-tahapan metode penelitian yang digunakan dalam penelitian. Dalam bab ini diuraikanlah tahap-tahapan untuk meneliti beserta penjelasannya.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan hasil dan pembahasan yang diharapkan dapat menyelesaikan permasalahan. Hasil dari penelitian akan diuraikan secara rinci pada bab ini.

BAB V PENUTUP

Bab ini berisikan kesimpulan dari hasil penelitian dan saran yang dapat digunakan untuk penelitian selanjutnya.

