

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Penggunaan teknologi dalam sektor pendidikan masih kurang dimanfaatkan dengan baik, salah satu teknologi yang bisa dimanfaatkan untuk sektor pendidikan adalah *Natural Language Processing* (NLP). Contoh penggunaan NLP dalam sektor pendidikan yaitu *Auto Essay Scoring* (AES). Tujuan dari AES adalah untuk memberikan skor pada jawaban siswa yang berbentuk essay secara otomatis berdasarkan kata yang ditulis siswa pada jawaban tersebut (Rajagede, 2021).

AES sudah digunakan pada beberapa penelitian, pada tahun (2018) menggunakan metode *Latent Semantic Analysis* (LSA) dan fitur N-gram (Citawan et al., 2018). Selanjutnya tahun (2020) peneliti menambahkan TF-IDF kedalam algoritma *Rabin Karp* untuk meningkatkan akurasi (Bahri & Wajhillah, 2020). Tahun (2021) menggunakan metode *BERT sentence embedding* (SBERT) dan *Neural Network* yang bertujuan untuk meningkatkan akurasi dari penelitian sebelumnya (Rajagede, 2021).

Pada tahun (2021), peneliti menggunakan *Feature Extraction* TF-IDF dan LSA dengan menggunakan 1648 jawaban essay yang telah di *preprocessing*. Penelitian ini menggunakan teknik klasifikasi dan algoritma similaritas dan menggunakan *Root Mean Square Error* (RMSE) untuk mengukur tingkat kesalahannya (Thamrin et al., 2021).

Selanjutnya tahun (2021) melanjutkan penelitian yang dilakukan Thamrin (2021) dengan data yang sama, peneliti menggunakan metode *regression* dan *oversampling* yang bertujuan untuk mengurangi skala error menggunakan *Support Vector Regression* (SVR), *Logistic Regression* (LR), dan *MLP Regression* (MLP-R). *Oversampling* menggunakan SMOTE untuk menyetarakan nilai training data lalu membandingkan hasilnya dengan *Root Mean Square Error* (RMSE) (Verdikha et al., 2021).

Metode SVR memiliki beberapa parameter kernel, yaitu kernel *linear*, kernel *polynomial*, dan kernel *Gaussian Radial Basis Function* (RBF). Pada penelitian ini penulis menggunakan parameter kernel *polynomial* karena pada penelitian Kavitha & Kumar (2018) terbukti bahwa penggunaan kernel *polynomial* mendapatkan nilai RMSE terbaik di antara kernel lain (Kavitha & Kumar, 2018).

Dari pembahasan tersebut, penulis akan melakukan penelitian untuk membandingkan hasil penelitian yang dilakukan Verdikha et al., (2021) menggunakan metode *Support Vector Regression* (SVR) dengan fitur ekstraksi TF-IDF dan parameter kernel *polynomial*. Hasil penelitian akan menggunakan *Root Mean Square Error* (RMSE) untuk mengukur tingkat kesalahannya.

## **1.2 Rumusan Masalah**

Rumusan masalah yang didapatkan dari pembahasan latar belakang adalah sebagai berikut :

1. Berapakah hasil nilai evaluasi RMSE pada AES berbahasa Indonesia menggunakan metode SVR dengan kernel *polynomial*?
2. Bagaimana perbandingan hasil evaluasi RMSE menggunakan metode SVR dengan kernel *polynomial* dengan penelitian sebelumnya yang dilakukan Verdikha et al., (2021).

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian yang dilakukan berdasarkan latar belakang yang sudah dibahas sebelumnya adalah :

1. Melakukan ekstraksi data dengan fitur ekstraksi TF-IDF
2. Melakukan pemisahan data yang sudah di ekstraksi menggunakan data split
3. Melakukan regresi dengan menggunakan metode SVR dengan kernel *polynomial*.

4. Mencari nilai evaluasi RMSE dari penggunaan kernel *polynomial* pada AES berbahasa Indonesia dan membandingkan hasil evaluasi RMSE dengan penelitian sebelumnya yang dilakukan Verdikha et al., (2021).