

ABSTRAK

KLASIFIKASI SOAL BERDASARKAN KATEGORI TOPIK MENGUNAKAN METODE *SUPPORT VECTOR MACHINES* DAN *NAIVE BAYES* (Studi Kasus: Mata Kuliah DWBI S1 Sistem Informasi)

Oleh :

CHIKA ENGGAR PUSPITA

1202174292

Klasifikasi pertanyaan merupakan sebuah model ilmu komputer, yang bertujuan untuk menganalisis pertanyaan serta dapat memberi label pada setiap pertanyaan berdasarkan kategori yang ada. Pertanyaan soal dapat dikumpulkan dari beberapa materi atau topik yang banyak dan berbeda. Dengan adanya klasifikasi soal kuis ini juga dapat membantu mahasiswa dalam pengambilan keputusan untuk menentukan jenis soal berdasarkan kategori topiknya. Oleh karena itu, peneliti bermaksud untuk membuat model klasifikasi pertanyaan soal kuis *Data Warehouse* dan *Business Intelligence* yang dapat dikelompokkan menjadi topik *Data Warehouse*, *Business Intelligence*, Data Analitik, dan Pengukuran Kinerja. Topik-topik tersebut diperoleh dari soal kuis pada mata kuliah *Data Warehouse* dan *Business Intelligence* (DWBI) pada program studi S1 Sistem Informasi Telkom University angkatan 2018.

Salah satu cara untuk mengatasi permasalahan ini yaitu dengan menggunakan pendekatan *machine learning*. Peneliti juga membahas cara menangani data yang tidak seimbang dengan menggunakan teknik *oversampling* dengan SMOTE dan validasi data menggunakan *K-Fold Cross-Validation* yang hasil akhirnya akan dilakukan perbandingan antara *dataset* menggunakan SMOTE dengan *dataset* yang tidak menggunakan SMOTE

Pada penelitian kali ini peneliti menggunakan perbandingan algoritma *machine learning* yaitu *Naïve Bayes* dan *Support Vector Machine* menggunakan metode SMOTE dan *Cross-Validation*. Hasil dari penelitian ini menunjukkan hasil akurasi yang terbaik dan sangat membantu. Hasil yang diperoleh pada metode *cross-validation* sebelum SMOTE menghasilkan tingkat akurasi sebesar 82.02% untuk hasil sesudah melalui tahap SMOTE sebesar 94.79 % pada algoritma *Naïve Bayes*, sedangkan pada algoritma *Support Vector Machine* menghasilkan akurasi pada proses sebelum SMOTE sebesar 81.39% dan untuk hasil sesudah melalui SMOTE sebesar 96.52%.

Kata kunci— **Klasifikasi Pertanyaan; Pembelajaran Mesin; *Naive Bayes*; *Support Vector Machine*; *Cross-Validation***