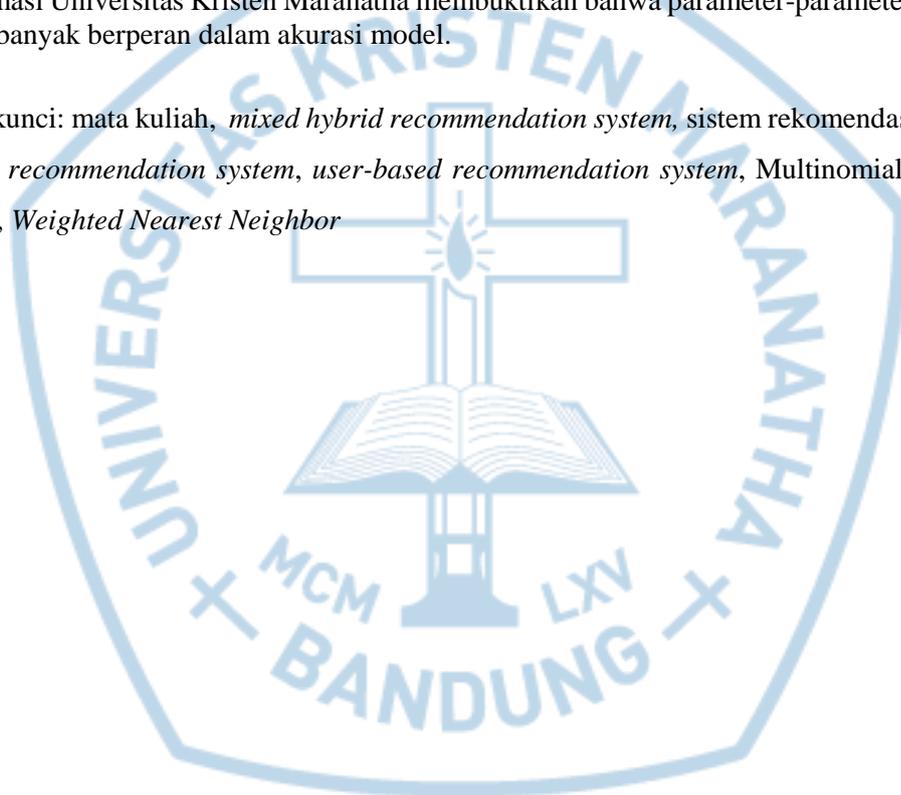


ABSTRAK

Perwalian sangat berperan penting dalam menentukan performa mahasiswa dalam masa pembelajaran di Universitas. Dosen wali sebagai pembimbing dari mahasiswa masih mengalami kesulitan dalam menentukan mata kuliah yang akan direkomendasikan kepada mahasiswa. Penelitian ini mengusulkan untuk mengembangkan sistem yang dapat memberikan rekomendasi mata kuliah untuk membantu dosen wali. Pada sistem yang akan dikembangkan dalam tugas akhir ini, pendekatan *mixed hybrid recommendation system: item-based* dan *user-based* filtering akan diterapkan. Berdasarkan hasil evaluasi yang dilakukan pada data terkontrol, sistem yang dikembangkan dapat memberikan rekomendasi dengan menggunakan model item-based pada tingkat akurasi 60% menggunakan algoritma Naive-bayes dan 58% untuk *user-based* menggunakan algoritma yang sama. McNemar test dapat membuktikan bahwa kedua model tersebut memiliki performa yang baik pada kasus yang berbeda. Pengujian model item-based terhadap data empiris Fakultas Teknologi Informasi Universitas Kristen Maranatha membuktikan bahwa parameter-parameter model tidak banyak berperan dalam akurasi model.

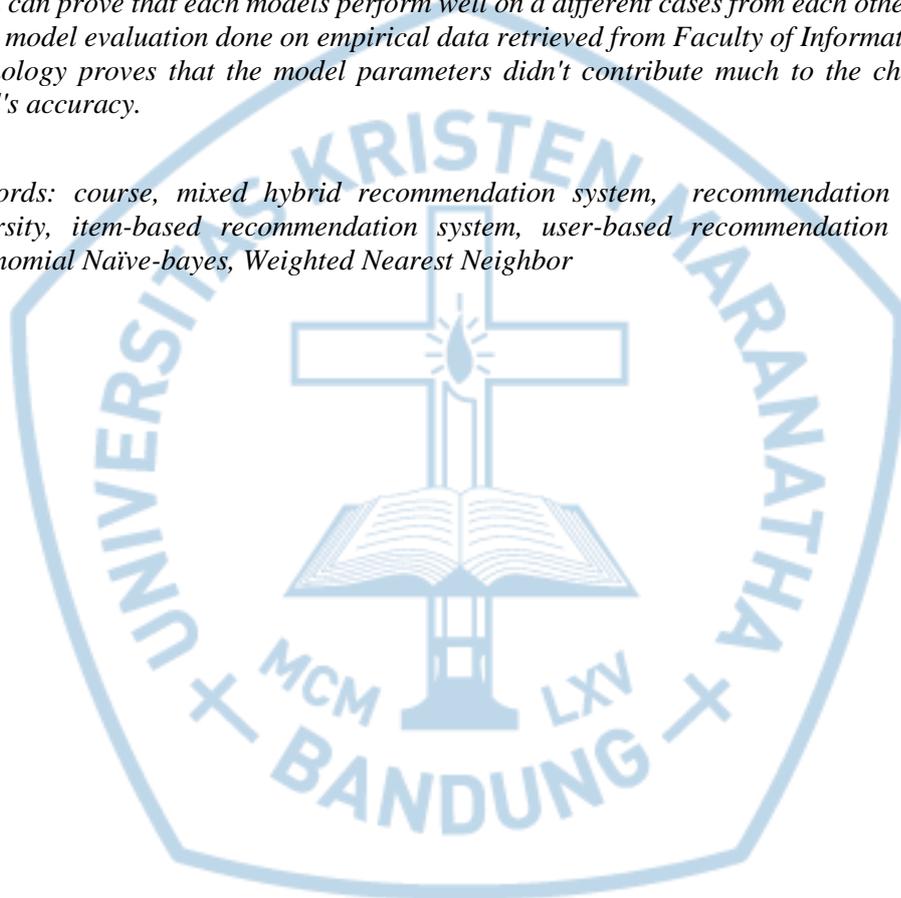
Kata kunci: mata kuliah, *mixed hybrid recommendation system*, sistem rekomendasi, *item-based recommendation system*, *user-based recommendation system*, Multinomial Naïve-bayes, *Weighted Nearest Neighbor*



ABSTRACT

Choosing the right courses is one of the most biggest factor that determine the performance of university students. The guidance lecturer as the one who supervise the students are having hard time to recommend the best courses for the students. This study propose a method to develop a course recommendation system to help the guidance lecturer. Those recommendations could be generated using machine learning methods. On the system that is developed in this study, we are using mixed hybrid recommendation system approach: item-based and user-based filtering. The evaluation result on controlled data show that the item-based model built using Naive-bayes algorithm land on 60% accuracy level and 58% accuracy on user-based model built using the same algorithm. McNemar test done on both model can prove that each models perform well on a different cases from each other. Item-based model evaluation done on empirical data retrieved from Faculty of Information and Technology proves that the model parameters didn't contribute much to the change of model's accuracy.

Keywords: course, mixed hybrid recommendation system, recommendation system, university, item-based recommendation system, user-based recommendation system, Multinomial Naïve-bayes, Weighted Nearest Neighbor



DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS LAPORAN PENELITIAN	ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN	iii
PRAKATA.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Pembahasan	2
1.4 Ruang Lingkup.....	2
1.5 Sumber Data.....	3
1.6 Sistematika Penyajian	3
BAB 2 KAJIAN TEORI	4
2.1 Sistem Rekomendasi	4
2.1.1 Pengaplikasian Sistem Rekomendasi.....	7
2.2 Sistem Klasifikasi	9
2.3 Naive-Bayes Classifier.....	10
2.3.1 Multinomial Naïve-Bayes	11
2.4 X^2 Feature Selection.....	11
2.5 Algoritma Distance-Weighted Nearest Neighbor	12

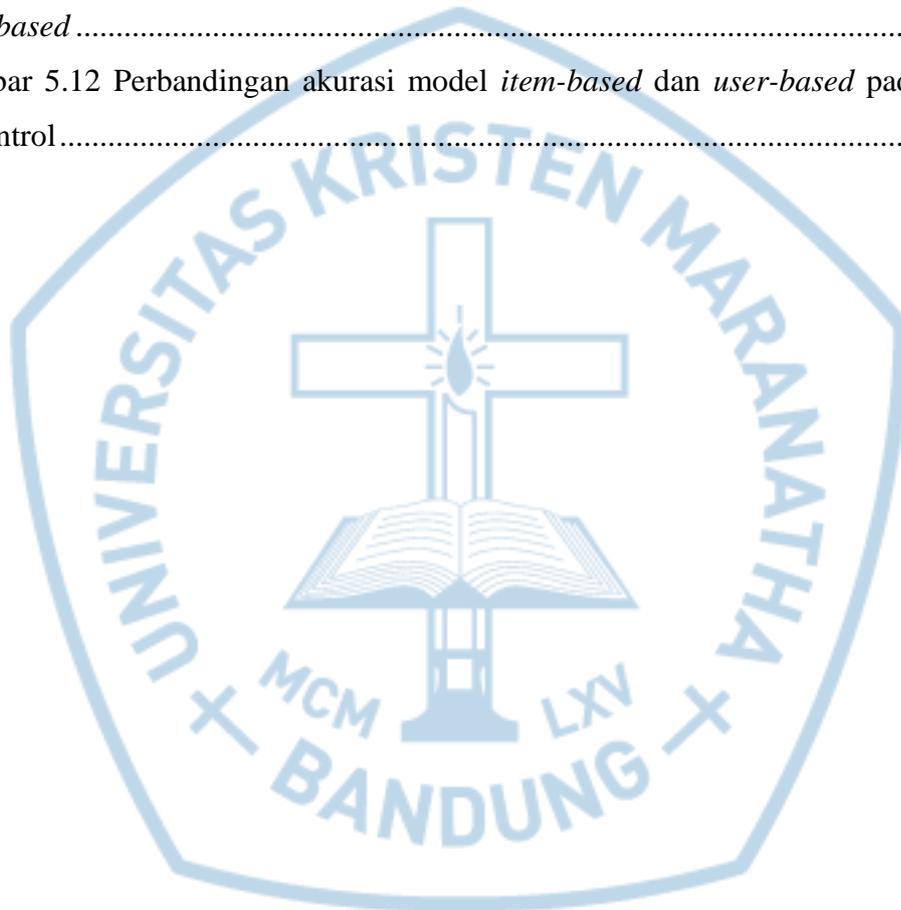
2.6 Scikit-learn	12
2.7 Python Sastrawi.....	13
2.8 SQLAlchemy	13
2.9 Flask.....	13
2.10 McNemar Test.....	13
2.11 K-fold <i>Cross-validation</i>	14
BAB 3 ANALISIS DAN RANCANGAN SISTEM.....	15
3.1 Sistem yang dikembangkan.....	15
3.2 Interaksi Sistem Dengan User.....	17
3.3 Desain Basis Data	20
3.4 Desain Tampilan Antar Muka.....	21
BAB 4 IMPLEMENTASI.....	23
4.1 Implementasi Tampilan Antar Muka	23
4.2 Implementasi Pengunggahan File Training	25
4.3 Instalasi Scikit-learn.....	26
4.4 X^2 <i>Feature Selection</i> pada Scikit-learn.....	27
4.5 Multinomial Naïve-bayes pada Scikit-learn.....	27
4.6 Weighted Nearest Neighbor pada Scikit-Learn	28
BAB 5 PENGUJIAN	29
5.1 Pengujian Model <i>Item-based Recommendation</i>	29
5.1.1 Data Pengujian Model <i>Item-based Recommendation</i>	29
5.1.1.1 Transformasi Data Empiris	30
5.1.2 Skema Pengujian Model <i>Item-Based</i>	32
5.1.3 Sanity Test pada Penerapan <i>Stopwords</i> Model Item-Based.....	33
5.1.4 Pengujian Dampak <i>Stopwords</i> Terhadap Akurasi Model <i>Item-based</i> .	37

5.1.5 Pengujian Dampak Algoritma Klasifikasi Terhadap Akurasi Model <i>Item-based</i>	39
5.1.6 Pengujian Dampak n-gram Terhadap Akurasi Model <i>Item-based</i>	40
5.1.7 Pengujian Dampak Jumlah Fitur pada x^2 <i>feature selection</i> Terhadap Akurasi Model <i>Item-Based</i>	40
5.1.8 Pengujian Dampak Jenis File Mata Kuliah Terhadap Akurasi Model <i>Item-Based</i>	41
5.1.9 Pengujian Dampak Jumlah <i>Neighbor</i> Algoritma KNN Terhadap Akurasi Model <i>Item-Based</i>	43
5.1.10 Pengujian Akurasi Tiap Proyek Pendidikan Pada Model <i>Item-based</i>	43
5.1.11 Pengujian K-fold Untuk Melihat Akurasi Prediksi Per Mahasiswa...	44
5.2 Pengujian Model <i>User-based</i>	45
5.2.1 Data yang Digunakan Dalam Pengujian Model <i>User-Based</i>	45
5.2.2 Skema Pengujian Model <i>User-Based</i>	46
5.2.3 Perbandingan Akurasi Algoritma KNN dan Naive-Bayes Untuk Model <i>User-based</i>	46
5.2.4 Pengujian dampak jumlah <i>neighbor</i> Algoritma KNN terhadap akurasi model <i>User-based</i>	47
5.3 Pengujian Tingkat Akurasi Model <i>Item-Based</i> dan <i>User-Based</i>	48
BAB 6 SIMPULAN DAN SARAN	50
6.1 Simpulan	50
6.2 Saran.....	51
DAFTAR PUSTAKA	1

DAFTAR GAMBAR

Gambar 2.1 <i>User-based vs Item-based</i> [13].....	5
Gambar 3.1 Model K-nearest Neighbor.....	17
Gambar 3.2 Use Case Diagram.....	18
Gambar 3.3 Activity Diagram untuk pengunggahan file.....	19
Gambar 3.4 Activity Diagram Lihat Rekomendasi.....	20
Gambar 3.5 Desain Basis Data.....	20
Gambar 3.7 Desain tampilan saat memilih mahasiswa.....	21
Gambar 3.8 Desain tampilan saat memilih mata kuliah yang akan diprediksi.....	21
Gambar 3.9 Desain tampilan hasil rekomendasi.....	22
Gambar 3.9 Desain tampilan untuk uploading file.....	22
Gambar 4.1 Implementasi tampilan awal sistem.....	23
Gambar 4.2 Implementasi tampilan memilih mata kuliah.....	24
Gambar 4.3 Implementasi tampilan hasil rekomendasi.....	24
Gambar 4.4 Implementasi tampilan unggah file.....	25
Gambar 4.5 Folder penyimpanan file mata kuliah.....	25
Gambar 4.6 Contoh hasil konversi file pdf.....	26
Gambar 4.7 Kode Program CountVectorizer.....	27
Gambar 4.8 Kode Program Chi2.....	27
Gambar 4.9 Kode Program Multinomial NB.....	27
Gambar 4.10 Kode Program Testing Multinomial Naïve-bayes.....	28
Gambar 4.11 Kode Program Weighted Nearest Neighbor.....	28
Gambar 5.1 Skema pembagian data pada pengujian data empiris.....	33
Gambar 5.2 Hasil perbandingan <i>stopwords</i> yang digunakan terhadap akurasi model <i>item-based</i>	38
Gambar 5.3 Perbandingan akurasi algoritma Naive- Bayes dan KNN untuk model <i>item-based</i>	39
Gambar 5.4 Perbandingan <i>n-gram language model</i> terhadap akurasi model <i>item-based</i>	40
Gambar 5.5 Perbandingan jumlah fitur terhadap akurasi model <i>item-based</i>	41

Gambar 5.6 Hasil perbandingan file training model <i>item-based</i> terhadap akurasi model.....	42
Gambar 5.7 Hasil perbandingan jumlah neighbor terhadap akurasi model <i>item-based</i>	43
Gambar 5.8 Akurasi tiap proyek pendidikan untuk model <i>item-based</i>	44
Gambar 5.9 Hasil pengujian k-fold untuk akurasi data per mahasiswa.....	45
Gambar 5.10 Perbandingan akurasi algoritma untuk model <i>user-based</i>	47
Gambar 5.11 Hasil perbandingan jumlah neighbor KNN terhadap akurasi model <i>user-based</i>	48
Gambar 5.12 Perbandingan akurasi model <i>item-based</i> dan <i>user-based</i> pada data terkontrol.....	49



DAFTAR TABEL

Tabel 2.1 <i>Contingency Table</i>	14
Tabel 3.1 Contoh model Multinomial Naive-Bayes	16
Tabel 5.1 Deskripsi kolom pada data empiris	29
Tabel 5.2 Contoh baris data empiris	30
Tabel 5.3 Pemotongan data empiris	31
Tabel 5.4 Diskretisasi nilai pada data empiris	31
Tabel 5.5 Transformasi Proyek Pendidikan Data Empiris.....	31
Tabel 5.6 Contoh hasil transformasi data empiris & data terkontrol.....	32
Tabel 5.7 Hasil sanity test <i>feature selection</i> tanpa <i>stopwords</i>	34
Tabel 5.8 Hasil sanity test <i>feature selection</i> dengan <i>stopwords</i> indonesia dan inggris	35
Tabel 5.9 Hasil sanity test <i>feature selection</i> dengan <i>stopwords</i> indonesia, inggris dan peraturan tambahan	36
Tabel 5.10 Jumlah fitur representatif pengujian sanity test	37
Tabel 5.11 Contoh data terkontrol	46
Tabel 5.12 Contoh transformasi data terkontrol untuk model <i>item-based</i>	49