

USULAN TUGAS AKHIR

1. IDENTITAS PENGUSUL

Nama : AHMAD YUSUF
NRP : 5108100153
Dosen Wali : WAHYU SUADI, S. KOM., M.KOM.

2. JUDUL TUGAS AKHIR

***“Implementasi Spectral Clustering untuk Mendukung Model Bootstrap Regresi
Linear dalam Prediksi Nilai Mahasiswa”***

3. LATAR BELAKANG

Tujuan utama dari *educational data mining* (EDM) adalah untuk membangun model komputasional serta *tools* nya dalam menggali data yang terdapat dalam dunia pendidikan. Penggalan data di dunia pendidikan (*educational data mining*) merupakan hal penting karena penyelenggara pendidikan bisa mendapatkan informasi terkait dari data yang berjumlah besar dan beragam. Oleh karena itu diperlukan metode yang baik bukan hanya dalam mengolah data tetapi juga dalam memperoleh informasi yang tersimpan dalam data tersebut.

Salah satu aplikasi dari penggalan data dalam dunia pendidikan ini adalah prediksi nilai pada siswa. Nilai merupakan parameter dari penilaian sukses atau tidaknya siswa. Oleh karena itu, prediksi ini sangat penting dalam membantu penyelenggara pendidikan ataupun siswa dalam melakukan tindakan ke depan yang sesuai.

Model yang umumnya digunakan dalam prediksi adalah regresi linear. Dengan regresi akan didapatkan nilai prediksi sesuai dengan variabel-variabel yang mempengaruhi pada data. Hanya saja permasalahannya kualitas regresi biasanya tergantung pada model

training. Noise yang terlalu banyak pada data training dapat menyebabkan model regresi yang dibentuk kurang akurat, sehingga prediksi yang dihasilkan kurang baik.

Akurasi regresi yang kurang disebabkan oleh noise data training dapat diatasi dengan mengelompokkan (*clustering*) terlebih dahulu. Dengan adanya *clustering* data akan dikelompokkan berdasarkan kesamaan data sehingga noise pada data training dapat diatasi, harapannya akurasi dari model regresi yang dihasilkan lebih baik.

Penelitian terhadap penggunaan metode clustering dalam meningkatkan akurasi dari model prediksi pernah diusulkan oleh beberapa orang, Shubhendu Trivedi, Zachary A. Pardos, Gabor N. Sarkozy, Neil T. Heffernan. Metode yang dikembangkan menggunakan spectral clustering untuk menghasilkan model regresi linear secara lebih akurat^[1].

Tugas akhir ini bertujuan untuk mengimplementasikan metode spectral clustering yang mendukung bootstrap regresi linear dalam memprediksi nilai mahasiswa. Dengan menggunakan metode tersebut diharapkan prediksi nilai mahasiswa yang dihasilkan memiliki akurasi yang lebih baik sehingga dapat membantu penyelenggara pendidikan serta mahasiswa dalam mengambil keputusan langkah ke depannya.

4. TUJUAN TUGAS AKHIR

Tugas akhir ini bertujuan mengimplementasikan metode spectral clustering dan bootstrap regresi linear dalam memprediksi nilai mahasiswa.

5. MANFAAT TUGAS AKHIR

Harapannya hasil implementasi pada tugas akhir ini dapat dimanfaatkan oleh penyelenggara pendidikan untuk memprediksi nilai dari mahasiswanya. Dan nantinya hasil prediksi tersebut dapat membantu dalam pengambilan tindakan yang sesuai kedepannya.

6. RUMUSAN MASALAH

Rumusan masalah yang diangkat dalam Tugas Akhir ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan algoritma *spectral clustering*.
2. Bagaimana mengimplementasikan metode bootstrap regresi linear yang memanfaatkan spectral clustering.

3. Bagaimana memprediksi nilai mahasiswa dengan metode bootstrap regresi linear dan *spectral clustering*.

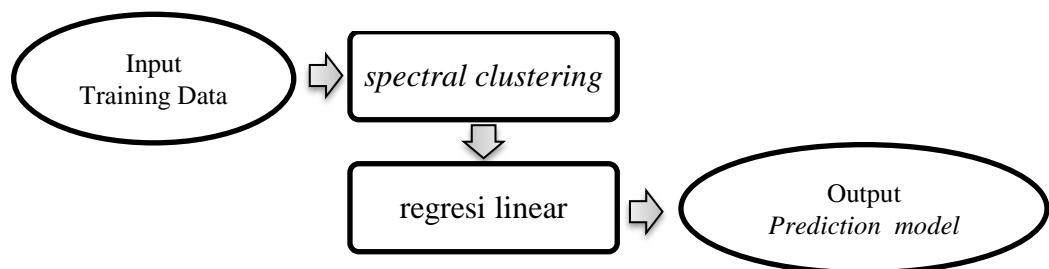
7. BATASAN MASALAH

Permasalahan yang dibahas dalam Tugas Akhir ini memiliki beberapa batasan, yaitu sebagai berikut :

1. Bahasa pemrograman yang digunakan dalam implementasi algoritma adalah *Java*
2. Pembentukan model pengelompokan menggunakan *spectral clustering* yang didukung dengan *linear regression*.
3. Permasalahan yang diujikan adalah memprediksikan performa mahasiswa dalam ujian.
4. Dataset mahasiswa berupa nilai tes dari mahasiswa yang bisa berasal dari Teknik Informatika ITS.

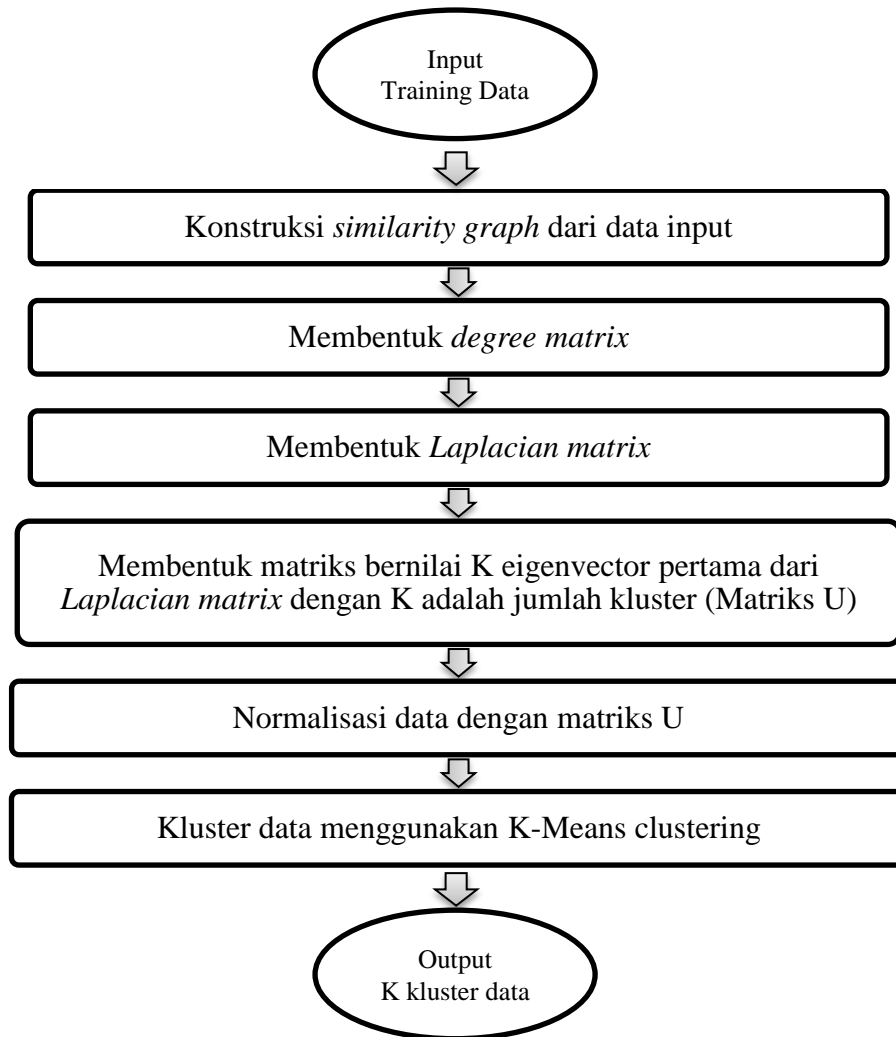
8. RINGKASAN TUGAS AKHIR

Langkah-langkah dalam memprediksi nilai mahasiswa dari dataset yang ada adalah sebagai berikut :



Dalam pembentukan *prediction model*, input yang digunakan adalah data training yang kemudian di kluster menjadi K kluster menggunakan algoritma *spectral clustering*. Setelah itu, pada setiap hasil klusternya dilakukan *linear regression* sehingga akan terbentuk suatu model persamaan pada setiap klusternya yang dinamakan *cluster model*. *Cluster model* nantinya akan membentuk suatu *prediction model* dapat digunakan dalam memprediksi suatu nilai keputusan dari data yang ada.

Langkah-langkah *spectral clustering* sendiri dapat digambarkan sebagai berikut :



Penjelasan dari langkah-langkah dalam melakukan *spectral clustering* adalah sebagai berikut :

1. Dari data training dibentuk *graph similarity* dimana n vertex pada graph tersebut merupakan representasi dari setiap record pada data training. Kemudian hubungkan setiap *vertex* ke k *vertex* terdekat lainnya dimana estimasi yang baik untuk nilai k adalah $\log(n)$. Hitung bobot dari setiap *vertex* yang terhubung (*edge*) dengan rumus :

$$w_{ij} = \exp\left(\frac{-\|s_i - s_j\|^2}{2\sigma^2}\right)$$

Dan dapat dibentuk matriks W yang berisi bobot dari setiap *edge* yang ada.

2. Dari *graph similarity* kemudian dihitung derajat dari setiap *vertex* dengan menjumlahkan bobot dari *edge* yang terhubung pada *vertex* yang bersangkutan atau dapat dirumuskan :

$$d_i = \sum_{j=1}^n w_{ij}$$

Dari derajat *vertex* tersebut dapat dibentuk *degree matrix* D yang merupakan matriks diagonal yang berisi bobot setiap *vertex*.

$$D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix}$$

3. Bentuk normalisasi *Laplacian matrix* dengan matriks W dan matriks D yang telah dihitung sebelumnya dengan rumus :

$$L_{sym} = D^{-1/2} W D^{-1/2}$$

4. Kemudian hitung K *eigenvector* pertama dari *Laplacian matrix* yang telah dihitung, dimana K merupakan jumlah kluster yang ingin ditemukan. Maka terbentuklah matriks $U = \{u_1, u_2, \dots, u_K\}$ yang merupakan K *eigenvector* pertama dari *Laplacian matrix*.
5. Normalisasi data dengan matriks U sehingga akan terbentuk K kolom yang merepresentasikan setiap nilai u pada setiap kolomnya.
6. Hasil dari data normalisasi kemudian dikluster dengan *K-means clustering* dimana jika data hasil normalisasi ke-i yang digunakan dalam perhitungan terkluster pada suatu kluster, maka data training ke i akan dimasukkan pada kluster tersebut.

Dari proses *spectral clustering* diatas akan dihasilkan data yang telah terkluster menjadi K kluster. Dari setiap klusternya dilakukan *linear regression* sehingga terdapat K persamaan regresi. Dan dari persamaan regresi tersebut akan dapat memprediksi dari data yang ada yaitu dengan menghitung persamaan regresi pada setiap kluster. Kemudian nilai prediksi dihasilkan dari rata-rata hasil perhitungan setiap persamaan regresi tersebut.

9. METODOLOGI

Metodologi yang akan dilakukan dalam Tugas Akhir ini memiliki beberapa tahapan, diantaranya sebagai berikut:

1. Penyusunan Proposal Tugas Akhir

Tahap awal untuk memulai pengerjaan Tugas Akhir adalah penyusunan Proposal Tugas Akhir. Pada proposal ini, penulis mengajukan gagasan pengimplementasian *spectral clustering* untuk mendukung model bootstrap regresi linear untuk prediksi nilai mahasiswa.

2. Studi Literatur

Pada tahap ini dilakukan pencarian, pengumpulan, penyaringan, pembelajaran dan pemahaman literatur yang berhubungan dengan data mining dan statistika, khususnya yang berkaitan dengan *clustering* dan regresi. Literatur yang digunakan dalam pengerjaan Tugas Akhir ini sebagian besar berasal dari internet berupa makalah ilmiah, tesis, artikel, materi kuliah, serta beberapa buku referensi.

3. Implementasi

Implementasi merupakan tahap dalam mengimplementasikan algoritma *spectral clustering*. Kemudian memodelkannya secara dinamis sehingga nantinya dapat memecahkan masalah *educational data mining* secara umum.

4. Pengujian dan Evaluasi

Pada tahap ini dilakukan uji coba terhadap model yang telah dibuat. Pengujian dilakukan dengan memprediksi performa mahasiswa dalam ujian dengan input dataset nilai mahasiswa. Sehingga kendala yang muncul dapat diidentifikasi serta ditemukan penyelesaiannya.

5. Penyusunan Buku Tugas Akhir

Tahap terakhir merupakan penyusunan laporan yang memuat dokumentasi mengenai pembuatan serta hasil dari implementasi perangkat lunak yang telah dibuat.

10. JADWAL PEMBUATAN TUGAS AKHIR

No	Tahapan	Bulan							
		September		Oktober		November		Desember	
1	Penyusunan Proposal								
2	Studi Literatur								
3	Implementasi								
4	Pengujian dan Evaluasi								
5	Penyusunan Buku Tugas Akhir								

11. DAFTAR PUSTAKA

- [1]. Shubhendu Trivedi, Zachary A. Pardos, Gabor N. Sarkozi, Neil T Hefferan. *Spectral Clustering in Data Mining*.
- [2].Ulrike von Luxburg. *A Tutorial on Spectral Clustering*.
- [3].Andrew Y. Ng, Michael I. Jordan, Yair Weiss. *On Spectral Clustering : Analysis and algorithm*.

LEMBAR PENGESAHAN

Surabaya, 15 Oktober 2011

Menyetujui,

Dosen Pembimbing I,

Dosen Pembimbing II,

(Dr. Ir. Raden Venantius Hari Ginardi, M. Sc.)
(NIP. 19650518 199203 1003)

(Isye Ariesianti, S.Kom., M.Phil.)
(NIP. 19780412 200604 2001)