

## **BAB II**

### **LANDASAN TEORI**

Bab ini membahas tentang teori-teori berupa tinjauan pustaka, tabel penelitian terdahulu serta ringkasan penelitian terdahulu.

#### **2.1 Educational Data Mining**

*Educational Data Mining* (EDM) merupakan metode untuk mengekstraksi informasi berharga yang berpotensi untuk mempengaruhi organisasi atau institusi pendidikan. Banyaknya penggunaan teknologi di dalam sistem pendidikan telah menyebabkan penyimpanan data dengan jumlah yang besar, sehingga dibutuhkan EDM untuk mengelola data tersebut. Beberapa manfaat dari EDM adalah untuk meningkatkan proses belajar mengajar, mengidentifikasi performa peserta didik, mengidentifikasi prioritas masalah dalam mendukung pembelajaran yang efisien, meningkatkan tingkat kelulusan, meningkatkan performa institusi pendidikan, memaksimalkan sumber daya institusi pendidikan, serta mengoptimalkan kurikulum mata pelajaran dalam institusi pendidikan [6]. Berikut ini adalah beberapa aplikasi utama dalam EDM menurut Romero dan Ventura [7]:

- 1) Menganalisis dan visualisasi data, bertujuan untuk mendukung keputusan dengan melakukan eksplorasi data.
- 2) Menyediakan umpan balik untuk mendukung pengajar, bertujuan untuk mendukung dan membantu pengajar dalam evaluasi diri yang lebih tepat.
- 3) Melakukan rekomendasi untuk peserta didik, bertujuan untuk memberikan rekomendasi langsung kepada peserta didik seperti rekomendasi pembelajaran, penyelesaian masalah, dan lain-lain.
- 4) Melakukan prediksi pada kinerja peserta didik, bertujuan untuk memperkirakan nilai variabel peserta didik yang tidak diketahui seperti pengetahuan, kinerja dan lain-lain.
- 5) Melakukan pemodelan terhadap peserta didik, bertujuan untuk mengembangkan model kognitif pengguna peserta didik seperti pemodelan karakteristik peserta didik (motivasi, kepuasan dan lain-lain).

- 6) Mendeteksi hal menyimpang dari perilaku peserta didik, bertujuan untuk memprediksi peserta didik yang memiliki beberapa jenis masalah, motivasi rendah, kecurangan, dan lain-lain.
- 7) Mengelompokkan peserta didik, bertujuan untuk mengelompokkan peserta didik sesuai karakteristik untuk mengembangkan sistem pembelajaran yang tepat.
- 8) Melakukan analisis terhadap jaringan sosial (*social network analysis*), bertujuan untuk mempelajari hubungan antar individu pada media sosial.
- 9) Membuat peta konsep, bertujuan untuk membantu pengajar dalam pembuatan peta konsep secara otomatis.
- 10) Melakukan konstruksi pada *Course Management System*, bertujuan supaya pengajar dapat membuat kurikulum dan mempelajari konten secara otomatis.
- 11) Melakukan perencanaan dan penjadwalan, bertujuan untuk meningkatkan proses pendidikan tradisional dengan otomatisasi seperti merencanakan pelajaran setiap semester, alokasi sumber daya dan lain-lain.

Pada penelitian ini, EDM diaplikasikan dengan melakukan analisis dan pengolahan data survei umpan balik kepada mahasiswa yang bertujuan untuk mendukung keputusan dengan melakukan eksplorasi data. Pada penelitian ini EDM dimanfaatkan untuk meningkatkan performa institusi pendidikan.

## **2.2 Tahapan Data Mining**

Tahapan *data mining* atau yang biasa disebut dengan *Knowledge Discovery from Data* (KDD) merupakan proses yang dilakukan untuk mendapatkan pola dan pengetahuan menarik dari data yang berjumlah besar [8]. Tahapan *data mining* yang dijalankan bertujuan untuk mendapatkan pengetahuan yang diinginkan. Penelitian ini melakukan proses *data mining* dengan mengikuti tujuh tahapan proses *data mining* yang dijabarkan menurut Jiawei Han, Micheline Kamber, dan Jian Pei. Berikut ini adalah tujuh tahapan dalam proses *data mining* menurut Jiawei Han, Micheline Kamber, dan Jian Pei [9]:

- 1) *Data Cleaning*, yaitu penghapusan data yang tidak konsisten dan penghapusan *noise* yang dapat mempengaruhi hasil dari proses *data mining*.
- 2) *Data Integration*, yaitu penggabungan data yang dari berbagai sumber penyimpanan data.
- 3) *Data Selection*, yaitu pemilihan data yang akan digunakan.
- 4) *Data Transformation*, yaitu pengubahan data sesuai dengan format yang sesuai dengan teknik *data mining* yang digunakan.
- 5) *Data Mining*, yaitu proses dalam menghasilkan pola dan model berdasarkan teknik *data mining* yang dipilih.
- 6) *Pattern Evaluation*, yaitu pencarian dan pengevaluasian pola serta model untuk mendapatkan pengetahuan yang diinginkan.
- 7) *Knowledge Presentation*, yaitu melakukan teknik visualisasi dan representasi pengetahuan untuk menyajikan pengetahuan yang dihasilkan kepada pengguna agar dapat lebih mudah dimengerti.

### **2.3 Open Ended Questionnaire & Close Ended Questionnaire**

*Open Ended Question* adalah bentuk pertanyaan atau pernyataan yang memberi kesempatan kepada responden untuk menuliskan pendapat mengenai pertanyaan atau pernyataan yang diberikan. Jawaban dari OEQ berbentuk kata-kata yang tidak memiliki skor numerik. OEQ mengharuskan responden untuk merumuskan jawaban atau respons dengan kata-kata mereka sendiri dan mengekspresikannya secara tertulis [10]. Responden tidak diarahkan ke arah tertentu sehingga dapat secara bebas menuliskan jawaban.

*Close Ended Question* (CEQ), yaitu bentuk pertanyaan atau pernyataan yang alternatif atau opsi jawabannya telah diberikan sehingga responden hanya dapat menjawab sesuai dengan pilihan yang ada. CEQ bertolak belakang dengan OEQ yang jawabannya berbentuk kata-kata secara bebas. Biasanya CEQ memiliki bentuk jawaban yang diberi dua opsi atau lebih. Jawaban tersebut dapat menunjukkan pilihan biner seperti benar atau salah, ataupun menunjukkan pilihan tingkat kesetujuan.

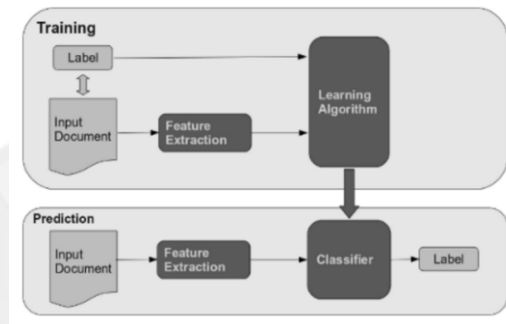
Hasil pengolahan dan analisis jawaban dari kedua bentuk kuesioner tersebut adalah salah satu alat yang umum digunakan institusi pendidikan untuk meningkatkan performa institusi. Melalui kuesioner, institusi pendidikan dapat mengetahui pendapat setiap pemangku kepentingannya [11]. Penelitian ini memanfaatkan EDM untuk mengolah survei umpan balik pada universitas yang berbentuk *Open Ended Question* (OEQ) dan *Close Ended Question* (CEQ). Contoh OEQ yang digunakan pada penelitian ini adalah “Apa saran Anda terkait proses belajar mengajar pada kelas ini?”. Contoh CEQ yang digunakan pada penelitian ini adalah “Dosen menerangkan pelajaran dengan jelas” dengan pilihan jawaban yaitu: (1) Sangat tidak setuju, (2) Tidak setuju, (3) Sedikit tidak setuju, (4) Cenderung setuju, (5) Setuju, dan (6) Sangat setuju.

## 2.4 Classification

*Classification* adalah teknik *data mining* yang digunakan untuk menganalisis data yang menghasilkan sebuah model untuk menjelaskan label data yang penting [8]. Tujuan utama dari *classification* adalah untuk mengidentifikasi kategori atau kelas terhadap suatu data. Proses *classification* secara umum digambarkan pada Gambar 2.1 [12]. Pada tahap pertama, data diberi label untuk digunakan sebagai masukan dalam tahap pembelajaran atau *training* yang kemudian dianalisis menggunakan algoritma klasifikasi. Tahap ini akan menghasilkan model atau *classifier*. Kemudian dilakukan tahap pengujian dengan menguji data yang belum memiliki label untuk memperoleh pengelompokan atau label [8] [12]. Data pada teknik *classification* diberikan label terlebih dahulu pada *training data*, maka dari itu teknik *classification* sering disebut sebagai *supervised learning*. *Supervised learning* adalah pembelajaran model *data mining* yang *data training*-nya telah diberi label, sehingga *output* dari *supervised learning* merupakan hasil dari pembelajaran sebelumnya.

*Classification* merupakan teknik *data mining* yang paling sering digunakan dalam pemanfaatan EDM dalam mengolah data survei umpan balik [13]. Penelitian ini menggunakan teknik *classification* untuk menganalisis data survei umpan balik institusi pendidikan yang berbentuk *Open Ended Question* (OEQ). Teknik

*classification* digunakan pada umpan balik dalam bentuk OEQ karena umpan balik tersebut akan dikelompokkan pada aspek-aspek survei yang telah ditentukan sebelumnya. Penelitian ini menggunakan dua algoritma *classification* yaitu *Support Vector Machine* (SVM) dan *Multi Layer Perception* (MLP) [8] [14].

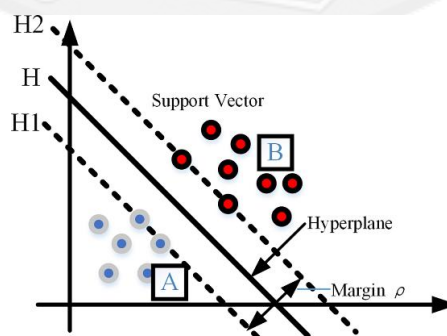


Gambar 2.1 Proses *Classification*

Sumber: J. George, S. N and S. George, "Classification Problem In Text Mining," *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, vol. 1, no. 8, 2014.

#### 2.4.1 *Support Vector Machine* (SVM)

*Support Vector Machine* (SVM) merupakan salah satu algoritma yang digunakan dalam *supervised learning*. Dasar-dasar SVM telah dikembangkan oleh Vapnik [15] dan mendapatkan popularitas di bidang *machine learning*. Prinsip dasar yang digunakan pada SVM adalah *linear classifier*. SVM dapat digunakan pada permasalahan non-linear dengan adanya konsep *kernel trick*. *Kernel trick* adalah metode sederhana dimana data non-linear dilakukan diproyeksikan ke ruang dimensi yang lebih tinggi sehingga lebih mudah untuk melakukan *classification* secara linear. Konsep *kernel trick* digunakan untuk dimensi yang lebih tinggi.



Gambar 2.2 *Support Vector Machine*

Sumber: Z. Deng, M. Cao, L. Rai and W. Gao, "A Two-Stage Classification Method for Borehole-Wall Images with Support Vector Machine," *PLoS ONE*, vol. 13, no. 6, 2018.



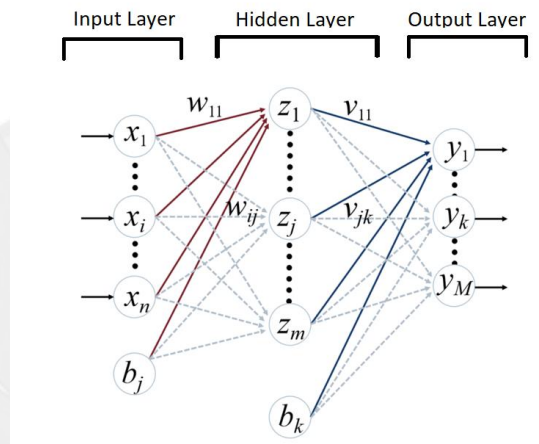
Gambar 2.2 menunjukkan pemisahan kelas yang berbeda dengan *hyperplane* yaitu garis pembatas antar kelas yang berbeda [16]. Objek yang digambarkan dengan A dan B pada Gambar 2.2 dinamakan *support vector* yang merupakan dua kelas yang memiliki jarak terdekat namun berasal dari kelompok yang berbeda. Semakin besar batas margin antara *support vector* dan *hyperplane*, semakin besar pula keakuratan dalam memprediksi data. Kelebihan dari SVM salah satunya adalah tidak terpengaruh dengan dimensi data sehingga sangat cocok untuk menyelesaikan masalah yang terkait dengan dimensi yang besar pada *information retrieval* [17].

#### **2.4.2 Multi Layer Perceptron (MLP)**

*Multi Layer Perceptron* (MLP) merupakan algoritma *Neural Network* yang paling umum digunakan dalam teknik *supervised* [18]. *Neural network* adalah sistem komputerisasi yang bekerja berdasarkan jaringan biologis (saraf) seperti otak manusia. Algoritma *Neural Network* dapat secara progresif melakukan *training* meningkatkan kinerjanya sendiri [19]. MLP adalah varian dari model *perceptron* yang pertama kali dikembangkan oleh Rosenblatt pada tahun 1950 [20]. MLP menggunakan pendekatan *backpropagation* untuk melakukan *training*. MLP terdiri dari paling sedikit tiga layer, yaitu layer *input*, satu atau beberapa layer tersembunyi (*hidden layer*) yang pada layer ini akan dilakukan *activation function* dan layer *output*. Gambar 2.3 merupakan struktur *feed forward* dari MLP. MLP dijalankan secara *feed forward*, artinya pemrosesan data diteruskan dari layer *input* sampai layer *output* kemudian dilakukan *backpropagation*, namun tidak menimbulkan suatu *loop* atau pengulangan dari satu neuron ke neuron lain. Proses *training* MLP dijabarkan sebagai berikut [21].

- 1) *Forward pass* dilakukan dengan meneruskan *input* ke dalam model dengan mengalikan bobot ( $w$ ) dan menambahkan bias ( $b$ ). *Activation function* dilakukan pada tahap ini. *Activation function* yang umum digunakan, yaitu: *sigmoid function*, dan *tanh function*.
- 2) Dilakukan *backpropagation* dengan melakukan evaluasi *error* (*loss calculate*).

- 3) Proses *backpropagation* dilanjutkan dengan melakukan koreksi pada bobot ( $w$ ) berdasarkan kalkulasi *error* yang telah didapat pada tahap *loss calculate*. Proses ini dinamakan *gradient descent* dimana pembaruan bobot dilakukan secara berulang untuk meminimalkan *loss function*. koreksi dilakukan secara terus menerus.



Gambar 2.3 Struktur *Feed Forward* MLP

Sumber: Z. E. Mohamed, "Using the artificial neural networks for prediction and validating solar radiation," *Journal of the Egyptian Mathematical Society*, 2019.

## 2.5 Clustering

*Clustering* adalah teknik *unsupervised learning* yang digunakan untuk mengelompokkan data yang mirip ke dalam kelas atau kelompok yang tidak ditetapkan sebelumnya [22]. Data dimasukkan ke dalam kelompok yang sama berdasarkan konektivitas dan kedekatannya. Performa teknik *clustering* yang baik diukur dari kemampuannya untuk mengidentifikasi pola yang tersembunyi dan mengidentifikasi pola kesamaan antar data yang mirip serta membedakan data yang tidak mirip ke dalam kelompok yang berbeda [23].

Teknik *clustering* secara umum telah banyak digunakan untuk pemanfaatan EDM [24]. Penelitian ini menggunakan teknik *clustering* untuk mengolah data umpan balik yang berbentuk *Close Ended Question* (CEQ) dalam mendapatkan kelompok responden. *Clustering* digunakan ketika label atau kelompok belum diketahui sebelumnya, maka *clustering* disebut sebagai *unsupervised learning* [23]. *Unsupervised learning* adalah pembelajaran model *data mining* yang pembelajaran datanya dilakukan secara *real time* (*training* dilakukan langsung menggunakan data

*input*), sehingga data yang menjadi *input* akan dianalisis tanpa melibatkan pembelajaran sebelumnya. *Unsupervised learning* berbeda dengan *supervised learning*, karena data pada *unsupervised learning* tidak memiliki label yang telah ditentukan sebelumnya. Penelitian ini menggunakan dua algoritma *clustering* yaitu *K-Means Clustering* dan *Fuzzy C-Means* (FCM). Pada penelitian ini, penentuan jumlah *cluster* dilakukan menggunakan *Elbow method*.

### 2.5.1 K-Means Clustering

*K-Means Clustering* adalah salah satu algoritma metode analisis data dalam *data mining* yang termasuk dalam *unsupervised learning*. Algoritma ini pertama kali dipublikasikan pada tahun 1984 oleh Stuart Lloyd [25]. *K-Means Clustering* termasuk dalam analisis non hierarki yang mengusahakan partisi objek ke dalam *cluster* atau kelompok berdasarkan karakteristiknya. *K-Means Clustering* menetapkan setiap *item* ke dalam *cluster* dengan rata-rata titik *centroid* terdekat. *Centroid* adalah titik pusat untuk setiap *cluster*. Pada *K-Means Clustering* dilakukan penghitungan berulang untuk mengoptimalkan posisi *centroid* [5]. Gambar 2.4 menunjukkan ilustrasi sederhana dalam melakukan teknik *clustering* dengan algoritma *K-Means Clustering* [26]. Langkah-langkah dalam menerapkan algoritma *K-Means Clustering* sebagai berikut:

- 1) Memilih secara acak data untuk dijadikan pusat *cluster*. Menghitung jarak antar pusat *cluster* dengan *Euclidean Distance*. Rumus 2.1 adalah persamaan *Euclidean Distance* untuk menghitung jarak data terhadap pusat *cluster*.

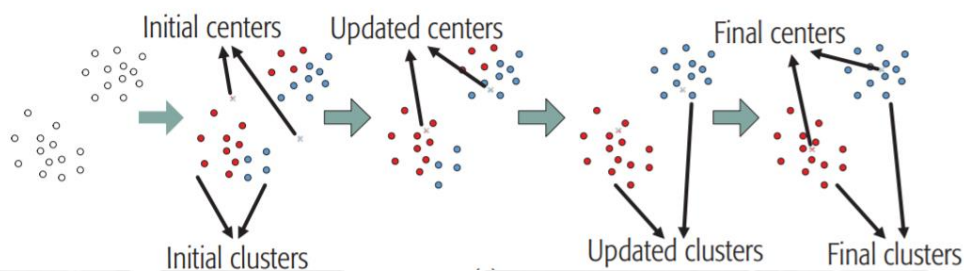
$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (2.1)$$

Keterangan:

- $D(i, j)$  = jarak data ke  $i$  terhadap pusat *cluster*  $j$
- $X_{ki}$  = data ke  $i$  pada atribut data ke  $k$
- $X_{kj}$  = titik pusat ke  $j$  pada atribut ke  $k$



- 2) Data ditempatkan di dalam *cluster* yang memiliki jarak terdekat, dihitung dari *centroid* terdekat.
- 3) Pusat *cluster* kemudian ditentukan jika seluruh data telah ditetapkan pada *cluster* terdekat.
- 4) Proses selanjutnya dilakukan iterasi penentuan pusat *cluster* dan penempatan data *cluster* sampai nilai *centroid* menjadi tetap atau tidak mengalami perubahan lagi [27].

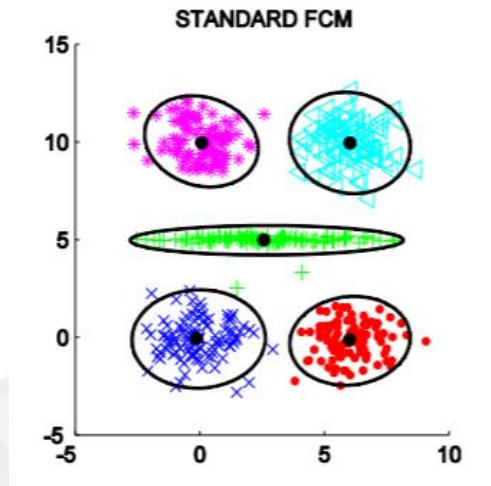


Gambar 2.4 *K-Means Clustering*

Sumber: R. He, B. Ai, A. F. Molisch, G. L. Stüber, Q. Li, Z. Zhong and J. Yu, "Clustering Enabled Wireless Channel Modeling Using Big Data Algorithms," *IEEE Communications Magazine*, 2018.

### 2.5.2 *Fuzzy C-Means (FCM)*

Algoritma *Fuzzy C-Means* (FCM) pertama kali diperkenalkan oleh Bazdek [28] pada tahun 1981 yang merupakan perluasan dari *Hard C-Mean clustering*. FCM merupakan algoritma yang termasuk dalam *unsupervised learning* yang diterapkan dalam beberapa masalah seperti klasifikasi, analisis fitur dan desain *classifier*. Algoritma FCM melakukan analisis dengan berbasis pada jarak antar titik data yang dimasukkan. *Cluster* dibentuk berdasarkan jarak antara titik data dengan pusat *cluster*. Pusat *cluster* dibentuk pada setiap *cluster*. FCM merupakan teknik pengelompokan data yang mengelompokkan kumpulan data menjadi  $n$  *cluster* [29]. Gambar 2.5 menunjukkan salah satu ilustrasi hasil *clustering* dengan algoritma FCM yang menghasilkan lima *cluster* [30].



Gambar 2.5 Fuzzy C-Means Clustering

Sumber: M. S. Yang and Y. Nataliani, "Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters," *Pattern Recognition*, 2017.

Algoritma FCM hampir sama dengan *K-Means Clustering* namun terdapat perbedaan yang utama. *K-Means Clustering* menempatkan elemen hanya pada satu *cluster* sedangkan FCM melakukan teknik *clustering* dengan metode *fuzzy* yaitu setiap elemen ditempatkan ke dalam *cluster* yang tersedia dengan perbedaan tingkat keanggotaan di setiap *cluster* tersebut. Kelebihan algoritma FCM salah satunya adalah kemampuan untuk mendeteksi *cluster* tingkat tinggi dan dapat ditunjukkan hubungan antar pola *cluster* yang berbeda [31].

### 2.5.3 Elbow Method

*Elbow Method* adalah metode yang digunakan untuk menentukan jumlah *cluster* yang optimal dalam pengembangan model *clustering* dengan metode visual [32]. *Elbow Method* menghasilkan *Elbow plot*, yaitu penggambaran garis antara pusat *cluster* dan nilai *distortion*. *Distortion* adalah jumlah kuadrat (*sum of square errors*) jarak dari suatu data poin pada setiap *cluster* ke titik tengah *cluster*-nya. Nilai *distortion* sama dengan 0 bila jumlah *cluster* sama dengan jumlah data yang menjadi *input* untuk dilakukan *clustering*.

*Elbow Method* dilakukan dengan menentukan kisaran jumlah *cluster*, yaitu  $k > 1$ . Dilakukan percobaan *clustering* dengan kisaran jumlah *cluster* yang telah ditentukan sehingga mendapatkan nilai *distortion* terhadap jumlah *cluster* yang

digambarkan dengan garis atau biasa disebut dengan *Elbow plot*. Jumlah *cluster* paling optimal saat garis membentuk lekukan atau sudut, yang sering disebut sebagai *Elbow criterion* [33].

## 2.6 Evaluasi Model

Dalam penggunaan teknik *classification*, sumber utama untuk mengestimasi akurasi kinerja sebuah model adalah *Confusion Matrix* atau *Classification Matrix* atau *Contingency Table* [34]. *Confusion Matrix* memberikan informasi dalam bentuk perbandingan antara hasil klasifikasi yang dilakukan oleh model dengan hasil klasifikasi sebenarnya. Terdapat empat istilah dalam pengukuran kinerja dengan menggunakan *Confusion Matrix* yaitu:

- 1) *True Positive* (TP) merupakan data positif yang diprediksi dengan benar. Kondisi ini dipenuhi apabila hasil klasifikasi secara aktual menunjukkan hasil yang benar (*true*) dan hasil klasifikasi oleh model juga menunjukkan hasil yang benar (*true*).
- 2) *True Negative* (TN) merupakan data negatif yang diprediksi dengan benar. Kondisi ini dipenuhi apabila hasil klasifikasi secara aktual menunjukkan hasil yang salah (*false*) dan hasil klasifikasi oleh model juga menunjukkan hasil yang salah (*false*).
- 3) *False Positive* (FP) atau *Type I Error* merupakan data negatif namun diprediksi sebagai data positif. Kondisi ini dipenuhi apabila hasil *classification* secara aktual menunjukkan hasil yang salah (*false*) dan hasil *classification* oleh model menunjukkan hasil yang benar (*true*).
- 4) *False Negative* (FN) atau *Type II Error* merupakan data positif namun diprediksi sebagai data negatif. Kondisi ini dipenuhi apabila hasil *classification* secara aktual menunjukkan hasil yang benar (*true*) dan hasil *classification* oleh model menunjukkan hasil yang salah (*false*).

Rumus 2.2, Rumus 2.3, Rumus 2.4, dan Rumus 2.5 adalah beberapa persamaan yang digunakan pada evaluasi kinerja model dengan *Confusion matrix* [34].

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2.2)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2.3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.4)$$

$$F_1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.5)$$

Keterangan:

- TP = *True Positive*
- TN = *True Negative*
- FP = *False Positive*
- FN = *False Negative*

Jika *classification* yang dilakukan melebihi dua kelas (*multi-class classification*), maka *Confusion Matrix* akan semakin besar (matriks kuadrat dengan jumlah label atau kelas yang unik), dan metrik akurasi menjadi semakin terbatas untuk setiap tingkat akurasi kelas (*Class Accuracy Rates*) dan seluruh akurasi *classifier* (*Overall Classifier Accuracy*).

Dalam mengevaluasi teknik *clustering*, biasanya tidak mudah karena tidak mempunyai kelompok data (k) sesungguhnya sebelum melakukan teknik *clustering* tersebut [35]. Jika sebelumnya telah diketahui kelompok data sebelum dilakukan teknik *clustering* (di mana cara ini tidak diterapkan dalam *clustering* yang aktual), maka dapat dilakukan pengevaluasian terhadap proses teknik *unsupervised* tersebut.

Evaluasi algoritma teknik *clustering* dapat dilakukan dengan *silhouette coefficient*. *Silhouette coefficient* dihitung dengan menganalisis jarak setiap data poin ke *cluster*-nya sendiri dan *cluster* lain. *Silhouette index* merupakan jarak rata-rata data poin ke seluruh data poin dalam *cluster*-nya sendiri dan jarak rata-rata data poin ke seluruh data poin pada *cluster* lain [36]. *Silhouette coefficient* menentukan apakah setiap data dalam satu *cluster* terkumpul dengan baik dan setiap data dalam *cluster* yang berbeda terpisah dengan baik. *Silhouette coefficient* yang bernilai 1, artinya *cluster* yang terbentuk oleh algoritma *clustering* tersebut terpisah dengan

sangat baik, dan *silhouette coefficient* yang bernilai -1 berarti terjadi kesalahan pada pembentukan *cluster*. Rumus 2.6 digunakan untuk menghitung *silhouette coefficient* untuk satu data poin ( $X^i$ ).

$$sil(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.6)$$

Keterangan:

- $sil(i)$  = nilai *silhouette coefficient* untuk poin X ke-i
- $a(i)$  = jarak poin X ke-i terhadap pusat *cluster*-nya
- $b(i)$  = jarak poin X ke-i terhadap titik *cluster* lain

Penelitian ini menggunakan *precision*, *recall*, *accuracy*, dan *F1-Score* untuk mengevaluasi model *classification*. Evaluasi model *clustering* dilakukan dengan menggunakan *silhouette coefficient*. Dalam penelitian ini, evaluasi pada model *classification* dilakukan untuk mengetahui kinerja setiap model dalam mengolah survei umpan balik dan memilih model yang terbaik untuk digunakan. Evaluasi pada model *clustering* dilakukan untuk mengetahui kinerja setiap model, yaitu perpaduan antara algoritma dan jumlah *cluster* yang ditentukan sehingga dapat memilih model terbaik untuk menghasilkan pengelompokan responden yang terbaik.

## 2.7 Visualisasi

Visualisasi adalah istilah untuk merepresentasikan data supaya dapat dikomunikasikan secara jelas dan efektif kepada pengguna dalam bentuk grafik. Visualisasi membantu pengguna untuk melihat hubungan data yang tidak mudah diobservasi jika dilihat dalam bentuk data mentah. Visualisasi memungkinkan kemudahan bagi pengguna dalam melakukan eksplorasi data karena bentuk representasi yang *user-friendly* [9]. Terdapat beberapa pendekatan umum dalam visualisasi data menurut Jiawei Han, Micheline Kamber, dan Jian Pei yaitu *Pixel-Oriented Visualization*, *Geometric Projection Visualization Techniques*, *Icon-based techniques*, dan *Hierarchical and Graph-Based Techniques*.






Penelitian ini melakukan visualisasi dengan bentuk grafik seperti *bar chart* dan *pie chart*. Selain itu penelitian ini juga melakukan visualisasi dalam bentuk *word cloud* dan tabel dua dimensi dengan fitur pencarian untuk mempermudah pengguna dalam melihat pengetahuan dari sekumpulan data dan informasi.

## 2.8 Business Process Model and Notation (BPMN)

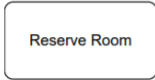
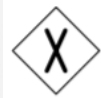

*Business Process Model and Notation* (BPMN) merupakan notasi standar untuk merepresentasikan atau memodelkan secara eksplisit sekumpulan proses dalam bentuk grafis [37]. BPMN diperkenalkan dan dikembangkan secara terkoordinasi dengan organisasi bernama *Object Management Group* (OMG) [38]. Tujuan BPMN adalah menyediakan gambaran yang mudah dipahami oleh semua pihak, mulai dari *business analysts* yang membuat rancangan proses bisnis, sampai pada pengembang teknis yang akan mengimplementasikan teknologi yang menjalankan proses tersebut, hingga pelaku bisnis yang akan mengelola dan memonitor proses tersebut [38]. Terdapat 4 kategori utama pada BPMN beserta elemen-elemennya, yaitu: *flow objects*, *artefacts*, *connecting objects*, dan *swimlanes* [38].

*Flow objects* terbagi menjadi tiga elemen yaitu *event*, *activity*, dan *gateway*. Tabel 2.1 menunjukkan penjelasan elemen dari *flow objects* beserta beberapa contoh dan pengertiannya [38].

Tabel 2.1 Contoh Elemen pada BPMN

Elemen	Penjelasan	Contoh Notasi	Fungsi
<i>Events</i>	<i>Events</i> digunakan untuk menghubungkan situasi dalam proses dengan proses yang berjalan. <i>Events</i> digunakan untuk menandakan jika ada situasi yang terjadi pada suatu proses.		<b>Start event:</b> menginisiasi jalannya sebuah proses
			<b>Intermediate event:</b> menunda proses, menunjukkan adanya sesuatu yang terjadi saat <i>event</i> berlangsung.
			<b>End event:</b> mengakhiri sebuah proses



Tabel 2.1 Contoh Elemen pada BPMN (lanjutan)

Elemen	Penjelasan	Contoh Notasi	Fungsi
<i>Activities</i>	<i>Activities</i> merupakan unit kerja yang merupakan unsur utama dalam proses bisnis. BPMN mendukung adanya <i>activities</i> yang memiliki hierarki atau memiliki sub-bagian atau cabang.		<b>Task:</b> menggambarkan suatu unit pekerjaan yang tidak dapat dipecah menjadi bisnis proses yang lebih detail.
<i>Gateways</i>	<i>Gateway</i> menggambarkan percabangan atau penggabungan alur.		<b>Exclusive (XOR):</b> menggambarkan alur alternatif yang hanya dapat diteruskan ke satu percabangan alur.
			<b>Parallel (AND):</b> menggambarkan alternatif alur yang secara paralel berjalan di setiap percabangan.

Sumber: M. Weske, Business Process Management: Concepts, Languages, Architectures, New York: Springer, 2007.

*Artefacts* digunakan untuk menunjukkan informasi tambahan tentang proses bisnis yang tidak secara langsung berhubungan dengan urutan alur bisnis proses seperti yang telah diatur oleh standar BPMN. Tabel 2.2 menunjukkan contoh elemen *artefacts* beserta fungsinya.




Tabel 2.2 Contoh Elemen *Artefacts* dan Fungsi

Elemen	Notasi	Fungsi
<i>Data object</i>		<i>Data object</i> menunjukkan data mana yang dibutuhkan untuk menjalankan suatu <i>activity</i> .
<i>Annotation</i>		<i>Annotation</i> digunakan untuk menunjukkan catatan agar membuat pembaca lebih memahami diagram bisnis proses.

Sumber: M. Weske, Business Process Management: Concepts, Languages, Architectures, New York: Springer, 2007.

*Connecting objects* berfungsi untuk menghubungkan *flow objects*, *swimlanes*, atau *artefacts*. *Connecting objects* terdiri dari *sequence flow*, *message flow*, dan *association*. Tabel 2.3 menunjukkan elemen-elemen *connecting objects* beserta fungsinya.

Tabel 2.3 Contoh Elemen *Connecting Objects* dan Fungsi

Elemen	Notasi	Fungsi
<i>Sequence Flow</i>		<i>Sequence Flow</i> digunakan untuk menentukan urutan alur obyek pada bisnis proses.
<i>Message Flow</i>		<i>Message Flow</i> berfungsi untuk menunjukkan aliran pesan antar pihak dalam bisnis proses yang direpresentasikan dengan <i>pool</i> .
<i>Association</i>		<i>Association</i> berfungsi untuk menghubungkan elemen pada bisnis proses dengan <i>artefacts</i> pada diagram proses bisnis.

Sumber: M. Weske, Business Process Management: Concepts, Languages, Architectures, New York: Springer, 2007.

*Swimlanes* adalah mekanisme visual yang mengelompokkan proses kegiatan yang terjadi dalam bisnis proses. *Swimlanes* terdiri dari *pool* dan *lane*. *Pool* merepresentasikan pihak utama dalam proses, biasanya satu *pool* mewakili sebuah departemen dalam perusahaan. Setiap *pool* memiliki satu atau lebih *lane*. *Lane* digunakan untuk melakukan kategorisasi *activities* berdasarkan fungsi atau perannya. Suatu *lane* dapat memiliki *flow objects*, *connecting objects*, dan *artifacts*. Gambar 2.6 menunjukkan notasi dari *swimlane* yang terdiri dari *pool* dan *lane*.

**Swimlanes**



Gambar 2.6 Notasi *Swimlane*

Sumber: M. Weske, Business Process Management: Concepts, Languages, Architectures, New York: Springer, 2007.

BPMN memberikan gambaran dalam proses yang dapat dijadikan dokumentasi yang perlu atau dapat dipahami oleh seluruh pihak yang terkait dalam proses [38]. Pada penelitian ini, BPMN digunakan untuk menggambarkan proses dalam pengembangan model *classification* dan model *clustering*. Selain itu BPMN

juga akan digunakan untuk menggambarkan proses penggunaan atau pemanfaatan model *classification*.

## 2.9 Penelitian Terdahulu

Tabel 2.4 dan Tabel 2.5 adalah tabel perbandingan penelitian terdahulu yang dirangkum berdasarkan relevansi dengan penelitian ini. Tabel 2.4 berfokus pada penelitian yang menggunakan teknik *Classification* untuk mengolah survei umpan balik dalam institusi pendidikan. Tabel 2.5 menunjukkan penelitian-penelitian sebelumnya yang menggunakan Teknik *Clustering* untuk mengolah survei umpan balik.

Tabel 2.4 terdiri dari 5 penelitian yaitu penelitian dengan judul “*Text Analytics Approach to Extract Course Improvement Suggestions from Students’ Feedback*” oleh Swapna Gottipati, Venky Shankararaman, dan Jeff Rongsheng Lin [39], “*Mining Student Feedback to Improve the Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic*” oleh Calandra Alencia Haryani, Achmad Nizar Hidayanto, Nur Fitriah Ayuning Budi, Zaenal Abidin, dan Theresia Wati [8], “*Aspect Based Opinion Mining On Student’s Feedback For Faculty Teaching Performance Evaluation*” oleh Irum Sindhu, Sher Muhammad Daudpota, Kamal Badar, Maheen Bakhtyar, Junaid Baber, dan Mohammad Nurunnabi [40], “*Text Mining in Survey Data*” oleh Christine P. Chai [41], “*Using Data Mining to Predict Instructor Performance*” oleh Ahmed Mohamed Ahmed, Ahmet Rizaner, dan Ali Hakan Ulusoy [14].

Tabel 2.4 Penelitian Sebelumnya Bagian 1

Atribut	Jurnal 1	Jurnal 2	Jurnal 3	Jurnal 4	Jurnal 5
<b>Penulis</b>	Swapna Gottipati, Venky Shankararaman, dan Jeff Rongsheng Lin	Calandra Alencia Haryani, Achmad Nizar Hidayanto, Nur Fitriah Ayuning Budi, Zaenal Abidin, dan Theresia Wati	Irum Sindhu, Sher Muhammad Daudpota, Kamal Badar, Maheen Bakhtyar, Junaid Baber, dan Mohammad Nurunnabi	Christine P. Chai	Ahmed Mohamed Ahmed, Ahmet Rizaner, dan Ali Hakan Ulusoy
<b>Judul</b>	<i>Text Analytics Approach to Extract Course Improvement Suggestions from Students' Feedback</i> [39]	<i>Mining Student Feedback to Improve the Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic</i> [8]	<i>Aspect Based Opinion Mining On Student's Feedback For Faculty Teaching Performance Evaluation</i> [40]	<i>Text Mining in Survey Data</i> [41]	<i>Using Data Mining to Predict Instructor Performance</i> [14]
<b>Teknik EDM</b>	<i>Classification</i>	<i>Classification</i>	<i>Classification</i>	<i>Classification</i>	<i>Classification</i>
<b>Obyek Penelitian</b>	Survei umpan balik mahasiswa bagian <i>Open Ended Question</i> (OEQ)	Survei umpan balik mahasiswa bagian <i>Open Ended Question</i> (OEQ)	Survei umpan balik mahasiswa bagian <i>Open Ended Question</i> (OEQ)	Survei umpan balik bagian <i>Close Ended Question</i> (CEQ) & <i>Open Ended Question</i> (OEQ)	Survei umpan balik mahasiswa bagian <i>Close Ended Question</i> (CEQ)



Tabel 2.4 Penelitian Sebelumnya Bagian 1 (lanjutan)

Atribut	Jurnal 1	Jurnal 2	Jurnal 3	Jurnal 4	Jurnal 5
<b>Perbandingan</b>	Penelitian ini menggunakan Teknik <i>Classification</i> dengan menggunakan algoritma <i>Support Vector Machine</i> (SVM), <i>General Linear Model</i> (GLM), <i>Decision Tree</i> (C5.0), <i>Conditional Inference Tree</i> (Ctree) serta menggunakan <i>Rule-Based classifiers</i> untuk mengklasifikasi komentar atau umpan balik mahasiswa dalam bentuk OEQ.	Penelitian ini menerapkan Teknik <i>Classification</i> dengan menggunakan algoritma <i>Decision Tree</i> , <i>Naïve Bayes</i> , <i>K-Nearest Neighbour</i> dan <i>Support Vector Machine</i> (SVM) untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk OEQ.	Penelitian ini menerapkan Teknik <i>Classification</i> dengan menggunakan algoritma LSTM ( <i>Long short-term memory</i> ) yang termasuk dalam <i>Neural Network Algorithm</i> untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk OEQ.	Penelitian ini menerapkan Teknik <i>Classification</i> dengan menggunakan algoritma <i>Supervised Latent Dirichlet Allocation</i> (sLDA) untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk OEQ. Hasil analisis tersebut juga dibandingkan dengan jawaban CEQ.	Penelitian ini menerapkan Teknik <i>Classification</i> dengan menggunakan algoritma <i>J48 Decision Tree</i> , <i>Multi Layer Perceptron</i> (MLP), <i>Naïve Bayes</i> , and <i>Sequential Minimal Optimization</i> (SMO) untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk CEQ.
<b>Keterbatasan</b>	Penelitian ini tidak memberikan area spesifik yang menjadi <i>concern</i> untuk <i>improvement</i> . Penelitian ini tidak melakukan analisis dengan jenis pertanyaan CEQ untuk lebih dalam menganalisis hasil survei umpan balik.	Penelitian ini tidak menganalisis umpan balik dari bentuk survei CEQ.	Penelitian ini tidak memberikan <i>multi label</i> terhadap survei umpan balik sehingga dapat terjadi kesalahan klasifikasi. Penelitian ini tidak menganalisis umpan balik survei dalam bentuk CEQ.	Penelitian ini hanya menggunakan satu algoritma untuk menganalisis umpan balik dalam bentuk OEQ. Selain itu, Pertanyaan CEQ tidak dianalisis menggunakan teknik <i>data mining</i> .	Penelitian ini tidak menganalisis survei umpan balik dalam bentuk OEQ.

Tabel 2.4 Penelitian Sebelumnya Bagian 1 (lanjutan)

Atribut	Jurnal 1	Jurnal 2	Jurnal 3	Jurnal 4	Jurnal 5
<b>Hasil</b>	Hasil penelitian ini menunjukkan bahwa algoritma Teknik <i>Classification</i> yang paling baik diantara algoritma-algoritma lain adalah algoritma C5.0 ( <i>Decision Tree</i> ) dengan nilai <i>precision</i> , <i>recall</i> dan <i>F-Score</i> yang tertinggi.	Hasil penelitian ini menunjukkan teknik klasifikasi yang terbaik diterapkan dengan menggunakan <i>Support Vector Machine</i> (SVM) dan dikombinasikan dengan fitur TFIDF, <i>Unigram</i> dan <i>Bigram</i> . Perpaduan teknik tersebut menghasilkan pelabelan aspek yang terbaik untuk survei umpan balik mahasiswa.	Hasil penelitian membuktikan algoritma yang digunakan dalam penelitian telah memberikan akurasi yang baik dengan ekstraksi aspek sebesar 91% dan <i>sentiment polarity detection</i> sebesar 93%.	Hasil penelitian memberi kesimpulan bahwa penggunaan hanya satu jenis kuesioner saja tidak dapat menjamin keakuratan data. OEQ memberikan pengetahuan yang lebih bernilai dan terkadang dapat bertentangan dengan skala numerik pada jawaban CEQ.	Hasil penelitian ini menunjukkan bahwa algoritma <i>Multilayer Perceptron</i> (MLP) dan <i>Sequential Minimal Optimization</i> (SMO) memberikan tingkat akurasi yang paling tinggi dalam menganalisis performa pengajar.

Tabel 2.5 terdiri dari 5 penelitian yaitu penelitian dengan judul “*Identifying Hidden Patterns in Students’ Feedback through Cluster Analysis*” oleh Anwar Muhammad Abaidullah, Naseer Ahmed, dan Edriss Ali [5], “*Cluster Analysis for Learning Style of Vocational High School Student Using K-Means and Fuzzy C-Means (FCM)*” oleh Reza Andrea, Shinta Palupi, Siti Qomariah [25], “*Utilizing Cluster Analysis of Close-Ended Survey Responses to Select Participants for Qualitative Data Collection in Mixed Methods Studies*” oleh Katherine M. Ehlert, Marian Kennedy, Courtney June Faber, dan Lisa Clemence Benson [42], “*Improving Quality of Educational Processes Providing New Knowledge using Data Mining Techniques*” oleh Manolis Chalaris, Stefanos Gritzalis, Manolis Maragoudakis, Cleo Sgouropoulou, dan Anastasios Tsolakidis [43], “*Clustering Students’ Open-Ended Questionnaire Answers*” oleh Wilhelmiina Hämäläinen, Mike Joy, Florian Berger, dan Sami Huttunen [44].

Tabel 2.5 Penelitian Sebelumnya Bagian 2

Atribut	Jurnal 6	Jurnal 7	Jurnal 8	Jurnal 9	Jurnal 10
<b>Penulis</b>	Anwar Muhammad Abaidullah, Naseer Ahmed, dan Edriss Ali	Reza Andrea, Shinta Palupi, Siti Qomariah	Katherine M. Ehlert, Marian Kennedy, Courtney June Faber, dan Lisa Clemence Benson	Manolis Chalaris, Stefanos Gritzalis, Manolis Maragoudakis, Cleo Sgouropoulou, dan Anastasios Tsolakidis	Wilhelmiina Hämäläinen, Mike Joy, Florian Berger, dan Sami Huttunen
<b>Judul</b>	<i>Identifying Hidden Patterns in Students' Feedback through Cluster Analysis [5]</i>	<i>Cluster Analysis for Learning Style of Vocational High School Student Using K-Means and Fuzzy C-Means (FCM) [25]</i>	<i>Utilizing Cluster Analysis of Close-Ended Survey Responses to Select Participants for Qualitative Data Collection in Mixed Methods Studies [42]</i>	<i>Improving Quality of Educational Processes Providing New Knowledge using Data Mining Techniques [43]</i>	<i>Clustering Students' Open-Ended Questionnaire Answers [44]</i>
<b>Teknik EDM</b>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering &amp; Association</i>	<i>Clustering</i>
<b>Obyek Penelitian</b>	Survei umpan balik mahasiswa bagian <i>Close Ended Question (CEQ)</i>	Survei umpan balik bagian <i>Close Ended Question (CEQ)</i>	Survei umpan balik mahasiswa bagian <i>Close Ended Question (CEQ)</i>	Survei umpan balik mahasiswa bagian <i>Close Ended Question (CEQ)</i>	Survei umpan balik mahasiswa bagian <i>Open Ended Question (OEQ)</i>

Tabel 2.5 Penelitian Sebelumnya Bagian 2 (lanjutan)

Atribut	Jurnal 6	Jurnal 7	Jurnal 8	Jurnal 9	Jurnal 10
<b>Perbandingan</b>	Penelitian ini menerapkan Teknik <i>Clustering</i> dengan menggunakan algoritma <i>K-Means Clustering</i> untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk CEQ.	Penelitian ini menerapkan Teknik <i>Clustering</i> dengan menggunakan algoritma <i>K-Means Clustering</i> dan <i>Fuzzy C-Means</i> (FCM) untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk CEQ.	Penelitian ini menerapkan Teknik <i>Clustering</i> dengan menggunakan algoritma <i>K-Means Clustering</i> , <i>Ward's</i> dan <i>Complete Link</i> untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk CEQ.	Penelitian ini menerapkan Teknik <i>Clustering</i> dan <i>Association Rule Mining</i> dengan menggunakan algoritma <i>K-Means Clustering</i> dan <i>Corelation Analysis</i> serta <i>FP-Growth</i> untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk CEQ.	Penelitian ini menerapkan Teknik <i>Clustering</i> dengan menggunakan algoritma <i>K-Means Clustering</i> , <i>Affinity Propagation</i> , dan <i>Spectral Clustering</i> untuk menganalisis survei umpan balik dari mahasiswa dalam bentuk OEQ.
<b>Keterbatasan</b>	Penelitian ini hanya menggunakan satu algoritma saja. Penelitian ini juga tidak melakukan evaluasi terhadap algoritma yang digunakan. Selain itu penelitian ini hanya mengolah data survei CEQ.	Penelitian ini hanya mengolah satu jenis survei, yaitu survei CEQ. Penelitian ini tidak mengolah data survei OEQ	Penelitian ini tidak melakukan analisis terhadap survei umpan balik dalam bentuk OEQ. Selain itu, penelitian ini juga belum dapat menentukan gambaran keseluruhan pada populasi.	Penelitian ini tidak menerapkan analisis terhadap umpan balik dengan bentuk OEQ.	Penelitian ini tidak menerapkan analisis survei umpan balik dalam bentuk CEQ.

Tabel 2.5 Penelitian Sebelumnya Bagian 2 (lanjutan)

Atribut	Jurnal 6	Jurnal 7	Jurnal 8	Jurnal 9	Jurnal 10
<b>Hasil</b>	Hasil penelitian ini menunjukkan bahwa teknik <i>clustering</i> dengan algoritma <i>K-Means Clustering</i> merupakan algoritma yang dapat dengan sederhana menunjukkan pola dan pengetahuan untuk membuat keputusan dalam mengevaluasi dan melakukan restrukturisasi kembali kurikulum serta meningkatkan kualitas pembelajaran mahasiswa.	Hasil penelitian ini menunjukkan eksperimen dengan menggunakan dua algoritma secara bersamaan untuk membandingkan hasil yang didapatkan. Keduanya menghasilkan 3 <i>cluster</i> dengan hasil yang sama, sedangkan 1 <i>cluster</i> dengan hasil yang berbeda.	Hasil penelitian ini menunjukkan penggunaan algoritma <i>K-Means Clustering</i> menjadi penerapan yang paling <i>reliable</i> dengan mempertimbangkan jumlah kuadrat dalam kluster, jumlah kuadrat antar kluster, dan <i>Calinski-Harabasz index (CH Index)</i> .	Hasil penelitian menunjukkan bahwa penerapan teknik <i>data mining</i> yang dikombinasikan dapat memberikan hasil yang lebih maksimal. Penggunaan campuran teknik-teknik dalam penelitian ini juga lebih memperlihatkan pengetahuan tersembunyi yang diinginkan <i>stakeholder</i> untuk meningkatkan kualitas pendidikan	Hasil penelitian ini menunjukkan bahwa algoritma <i>K-Means Clustering</i> dan <i>Affinity Propagation</i> memberikan hasil yang lebih baik dilihat dari evaluasi dengan <i>purity</i> dan NMI



## 2.10 Ringkasan Penelitian Terdahulu

Penelitian pada Tabel 2.4 dilakukan dengan menerapkan teknik *classification*, sedangkan penelitian pada Tabel 2.5 dilakukan dengan menerapkan teknik *clustering*. Pada penelitian oleh Swapna Gottipati, Venky Shankararaman, dan Jeff Rongsheng Lin dengan judul “*Text Analytics Approach to Extract Course Improvement Suggestions from Students’ Feedback*” dilakukan evaluasi terhadap teknik *classification* yang digunakan dan melakukan eksplorasi *suggestion* atau saran dari survei umpan balik mahasiswa yang berbentuk *Open Ended Question* (OEQ). Penelitian ini juga menunjukkan bahwa algoritma *classification* C5.0 (*Decision Tree*) yang paling baik [39]. Penelitian oleh Calandra Alencia Haryani, Achmad Nizar Hidayanto, Nur Fitriah Ayuning Budi, Zaenal Abidin, dan Theresia Wati dengan judul “*Mining Student Feedback to Improve the Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic*” merupakan penelitian yang menjadi referensi utama pada tugas akhir ini. Tugas akhir ini juga merupakan lanjutan dari penelitian yang dilakukan oleh Calandra Alencia Haryani, Achmad Nizar Hidayanto, Nur Fitriah Ayuning Budi, Zaenal Abidin, dan Theresia Wati tersebut. Penelitian yang berjudul “*Mining Student Feedback to Improve the Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic*” ini mendapatkan label aspek *classification* dan hasil *sentiment analysis* serta tren topik pada survei umpan balik mahasiswa dalam bentuk *Open Ended Question* (OEQ). Penelitian ini juga menunjukkan bahwa algoritma *Support Vector Machine* (SVM) menjadi algoritma yang paling baik pada penelitian tersebut [8]. Penelitian oleh Irum Sindhu, Sher Muhammad Daudpota, Kamal Badar, Maheen Bakhtyar, Junaid Baber, dan Mohammad Nurunnabi dengan judul “*Aspect Based Opinion Mining On Student’s Feedback For Faculty Teaching Performance Evaluation*” dilakukan pengolahan survei umpan balik mahasiswa untuk mendapatkan aspek dan *sentiment* dalam pengolahannya. Penelitian ini hanya menggunakan satu teknik *classification* yaitu *Long short-term memory* (LSTM) [40]. Penelitian oleh Christine P. Chai dengan judul “*Text Mining in Survey Data*” dilakukan pengolahan data survei umpan balik tentang kepuasan pegawai dengan menggunakan algoritma *supervised Latent*

*Dirichet Allocation* (sLDA) menganalisis survei umpan balik dalam bentuk *Open Ended Question* (OEQ) yang merupakan alasan jawaban dari *Close Ended Question* (CEQ) dengan jawaban yang berkisar antara 1 – 10 (skala interval). Pada penelitian ini, skala jawaban *Close Ended Question* (CEQ) diolah jumlah respondennya untuk melihat hubungan antara hasil kedua bentuk pertanyaan tersebut [41]. Penelitian oleh Ahmed Mohamed Ahmed, Ahmet Rizer, dan Ali Hakan Ulusoy dengan judul “*Text Mining in Survey Data*” dilakukan prediksi terhadap performa dosen dengan mengolah data survei umpan balik dengan bentuk *Close Ended Question* (CEQ). Penelitian ini menunjukkan bahwa algoritma *Multi Layer Perceptron* (MLP) dan *Sequential Minimal Optimization* (SMO) memberikan akurasi yang terbaik [14]. Penelitian oleh Anwar Muhammad Abaidullah, Naseer Ahmed, dan Edris Ali dengan judul “*Identifying Hidden Patterns in Students’ Feedback through Cluster Analysis*” dilakukan pengolahan data survei umpan balik yang berbentuk *Close Ended Question* (CEQ) dengan menggunakan algoritma *K-Means Clustering*. Penelitian ini bertujuan untuk pengambilan keputusan yang efektif oleh institusi pendidikan dan meningkatkan kualitas pembelajaran bagi peserta didik [5]. Penelitian oleh Reza Andrea, Shinta Palupi, Siti Qomariah dengan judul “*Cluster Analysis for Learning Style of Vocational High School Student Using K-Means and Fuzzy C-Means (FCM)*” dilakukan pengolahan terhadap data survei umpan balik terkait tipe pembelajaran yang menjadi preferensi peserta didik. Penelitian ini menggunakan dua algoritma dan membandingkan hasil dari kedua algoritma tersebut [25]. Penelitian oleh Katherine M. Ehlert, Marian Kennedy, Courtney June Faber, dan Lisa Clemence Benson dengan judul “*Utilizing Cluster Analysis of Close-Ended Survey Responses to Select Participants for Qualitative Data Collection in Mixed Methods Studies*” dilakukan pengolahan data survei umpan balik dari kuesioner peserta didik untuk mendapatkan kelompok yang cocok untuk dipilih menjadi calon peserta wawancara. Penelitian ini membandingkan tiga algoritma dan mendapatkan *K-Means Clustering* sebagai algoritma yang terbaik untuk kasus ini [42]. Penelitian oleh Manolis Chalaris, Stefanos Gritzalis, Manolis Maragoudakis, Cleo Sgouropoulou dengan judul “*Improving Quality of Educational Processes Providing New Knowledge using Data Mining Techniques*”

bertujuan untuk melakukan peningkatan performa dengan menerapkan teknik *data mining* dalam proses edukasi dalam institusinya. Eksperimen dalam penelitian ini mengolah data yang diperoleh dari kuesioner dengan pertanyaan berbentuk *Close Ended Question* (CEQ) kepada mahasiswa yang dibagikan pada setiap departemen. Penelitian ini melakukan *clustering* dengan *K-Means Clustering*, setelah itu melakukan pengolahan data dengan *corelation matrix* dan menggunakan teknik *association rules mining* [43]. Penelitian oleh Wilhelmiina Hämäläinen, Mike Joy, Florian Berger, dan Sami Huttunen Anastasios Tsolakidis dengan judul “*Clustering Students’ Open-Ended Questionnaire Answers*” dilakukan teknik *clustering* terhadap kuesioner dalam bentuk *Close Ended Question* (CEQ) dengan tujuan untuk membandingkan teknik *clustering* yang terbaik untuk dipadukan dengan ekstraksi topik dengan algoritma HITS. Pada penelitian ini evaluasi agar dapat melakukan ekstraksi topik dengan dua bahasa [44].

Penelitian-penelitian sebelumnya telah melakukan pengolahan terhadap beberapa jenis survei umpan balik baik tertutup maupun terbuka. Namun pada penelitian-penelitian sebelumnya jarang dilakukan pengolahan dengan dua bentuk pertanyaan survei umpan balik. Oleh karena itu, pada penelitian ini dilakukan analisis terhadap dua bentuk survei umpan balik yaitu *Open Ended Question* (OEQ) dan *Close Ended Question* (CEQ) untuk memperoleh hasil yang lebih maksimal. Penelitian ini menggunakan dua teknik *data mining* yaitu teknik *classification* dan teknik *clustering*. Selain itu, pada penelitian ini juga dilakukan evaluasi terhadap setiap algoritma yang digunakan untuk memilih model yang terbaik dalam penelitian ini. Kemudian dilakukan visualisasi untuk mempermudah pengguna dalam mendapatkan *insight* yang lebih baik terhadap hasil penelitian.