

BAB II

TINJAUAN PUSTAKA

2.1 E-Commerce

E-commerce yang berarti perdagangan elektronik merupakan urusan dengan barang dan jasa melalui media elektronik dan *internet*. *e-commerce* melibatkan menjalankan bisnis dengan bantuan *internet* dan dengan menggunakan teknologi informasi seperti *Electronic Data Interchange* (EDI). *e-commerce* berhubungan dengan situs *web vendor* di *internet*, yang memperdagangkan produk atau layanan langsung ke pelanggan dari *portal*. *Portal* menggunakan aplikasi keranjang belanja digital dan memungkinkan pembayaran melalui kartu kredit, kartu debit atau pembayaran EFT (*Electronic fund transfer*) (Bhalekar, et al., 2014).

Menurut (Cashman, 2007) *e-commerce* atau kependekan dari elektronik *commerce* (perdagangan secara elektronik), merupakan transaksi bisnis yang terjadi dalam jaringan elektronik, seperti *internet*. Siapapun yang dapat mengakses komputer, memiliki sambungan ke *internet*, dan memiliki cara untuk membayar barang-barang atau jasa yang dibeli, dapat berpartisipasi dalam *e-commerce*.

Berdasarkan jenis hubungan antara berbagai *e-commerce*, dapat dikategorikan dalam berbagai jenis yaitu:

1. B2B (*Business-to-Business*)

Perusahaan melakukan bisnis satu sama lain seperti produsen menjual ke distributor dan grosir menjual ke *retailers*. Harga didasarkan pada jumlah pesanan dan dapat dinegosiasikan.

2. B2C (*Business-to-Consumer*)

Jenis bisnis yang dilakukan antara pelaku bisnis dengan konsumen, seperti antara produsen yang menjual dan menawarkan produknya ke konsumen umum secara *online*.

3. C2B (*Consumer-to-Business*)

Seorang konsumen memposting proyeknya dengan anggaran yang ditetapkan secara *online* dan dalam batas waktu tertentu, perusahaan meninjau persyaratan dan tawaran konsumen pada proyek tersebut. Konsumen meninjau tawaran dan memilih perusahaan yang akan menyelesaikan proyek. Memberdayakan konsumen di seluruh dunia dengan menyediakan tempat pertemuan dan *platform* untuk melakukan transaksi.

4. C2C (*Consumer-to-Consumer*)

C2C merupakan salah satu model *e-commerce* yang menjual produknya secara langsung antar konsumen satu ke konsumen lainnya, atau dapat dikatakan sebagai transaksi jual beli antar konsumen.

Beberapa keuntungan yang dapat diambil dengan adanya *e-commerce*, yaitu *revenue* (aliran pendapatan) baru yang mungkin lebih menjanjikan yang tidak bisa ditemui di aplikasi transaksi pasar tradisional tetapi, dapat meningkatkan *market exposure* (pangsa pasar), menurunkan biaya operasional (*operating cost*), melebarkan jangkauan *global reach*, meningkatkan loyalitas konsumen *consumer loyalty*, meningkatkan *supplier management*, memperpendek waktu produksi, dan meningkatkan mata rantai pendapatan (*value chain*) (Cashman, 2007).

Meskipun *e-commerce* merupakan aplikasi yang menguntungkan karena dapat mengurangi biaya transaksi bisnis dan dapat memperbaiki kualitas pelayanan, namun aplikasi *e-commerce* ini beserta semua infrastruktur pendukungnya mudah sekali untuk disalahgunakan oleh pihak-pihak yang tidak bertanggung jawab. Bisa jadi kesalahan-kesalahan yang mungkin timbul melalui berbagai cara kerusakan hebat yang terjadi pada semua element yang berkaitan dengan aplikasi. Dari segi pandangan bisnis, penyalahgunaan dari kegagalan aplikasi yang terjadi seperti, kehilangan dari segi financial secara langsung karena kecurangan, pencurian informasi rahasia yang berharga, kehilangan kesempatan bisnis karena gangguan pelayanan, penggunaan akses ke sumber data penting oleh pihak yang tidak memiliki hak, kehilangan kepercayaan dari para konsumen, dan kerugian-kerugian yang tak terduga (Cashman, 2007).



2.2 Transaksi *Online*

E-commerce mengacu sebagai segala bentuk transaksi dari bisnis yang dilakukan secara online dan sebagai salah satu bentuk transaksi *online* (Markus, 2019) yang merupakan suatu proses transaksi yang dilakukan dan memerlukan internet, proses transaksi *online* disebut sebagai *Online Transaction Processing* (OLTP), OLTP merupakan kelas dari program perangkat lunak yang mampu mendukung aplikasi yang sifatnya berorientasi pada transaksi (Rouse, 2013), menurut Ian OLTP merupakan sebuah kategori dari *data processing* yang berfokus pada tugas yang sifatnya berorientasi pada transaksi , dan OLTP secara khusus terlibat dalam *insert* , *update* dan *delete* data dalam jumlah kecil dalam sebuah *database*.

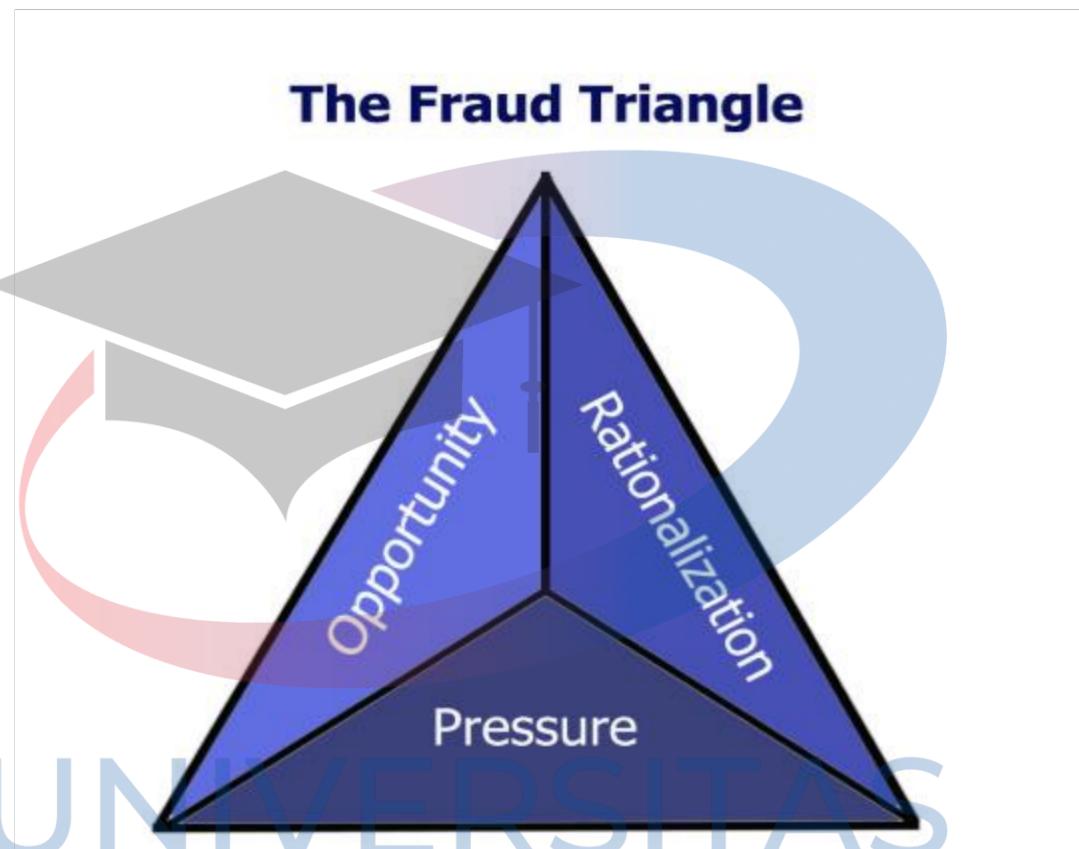
Karakteristik transaksi *online* atau OLTP biasanya sangatlah spesifik dalam tugas yang dilakukan, dan biasanya melibatkan *record* atau seleksi kecil dari *record*, karakteristik dari aplikasi *e-commerce* atau aplikasi OLTP adalah sebagai berikut:

1. Transaksi yang melibatkan data dalam jumlah yang kecil
2. Akses data yang sifatnya terindeks
3. Angka pengguna dalam jumlah yang besar
4. *Updates* dan *queries* yang sering terjadi
5. Waktu respon yang cepat

Aplikasi *e-commerce* secara khas harus memiliki ketersediaan yang sangatlah tinggi, dikarenakan aplikasi OLTP sering berhadapan dengan data yang sangat kritis, dan jumlah pengguna yang sangat banyak , ketersediaan yang buruk menyebabkan hilangnya pengguna dan berakibatkan pengguna baru tidak puas terhadap aplikasi *e-commerce* yang ada (IAN, 2017).

2.3 Penipuan

Menurut (Cressey, 1950) mengapa orang-orang melakukan kejahatan penipuan adalah karena faktor Fraud Triangle dimana Fraud Triangle merupakan suatu model untuk menjelaskan faktor-faktor yang menyebabkan seseorang untuk melakukannya tindakan penipuan.



Gambar 2.1 *The Fraud Triangle* (Cressey, 1950)

Menurut (Cressey, 1950), ada tiga faktor yang harus bersamaan terjadi agar seseorang yang normal untuk melakukan tindakan penipuan, ketiga faktor tersebut secara umum direferensikan sebagai The Fraud Triangle, termasuk konsep dari *Pressure* (Tekanan), *Opportunity* (Peluang) dan *Rationalization* (Rasionalisasi) dengan penjelasan sebagai berikut:

1. *Pressure* (Tekanan)

Tekanan adalah yang memotivasi seorang individu untuk melakukan tindak penipuan.

Tekanan yang termasuk hampir semuanya seperti pengeluaran medis yang tidak dapat dibayarkan, gaya hidup mewah, pencapaian pekerjaan, perjuangan dengan kecanduan dan

lain-lain. Tekanan bisa jadi pribadi atau professional, tetapi bisa juga berupa permasalahan keuangan yang signifikan dan tidak dapat diselesaikan dengan cara yang sah.

2. *Opportunity* (Peluang)

Peluang merupakan metode yang digunakan individu untuk melakukan tindak penipuan. Peluang merupakan ciri khas yang diciptakan oleh kontrol internal yang lemah, manajemen pengawasan yang buruk, penyalahgunaan wewenang, kegagalan penetapan prosedur yang baik untuk mendeteksi aktivitas penipuan juga sebagai peningkatan peluang agar penipuan terjadi, bagian ini yang merupakan bagian The Fraud Triangle adalah yang termudah untuk dikontrol dimana dari bagian ini dapat dibuatkan pengukuran yang dapat mencegah penipuan.

3. *Rationalization* (Rasionalisasi)

Rasionalisasi adalah pemberian alasan yang digunakan oleh individu untuk melakukan penipuan, individu mayoritas yang luas yang melakukan kejahatan adalah pelaku pertama, dan tidak melihat diri mereka sebagai kriminal – kriminal, melainkan orang- orang biasa yang jujur yang hanya menjadi korban dari keadaan yang tidak menguntungkan.

Kunci untuk menghalangi penipuan adalah melanggar Fraud Triangle ini. Beberapa dari jenis penipuan merupakan bagian dari penipuan *online*, penipuan *online* menurut (Team, 2019) beroperasi dalam banyak nama dan penyamaran termasuk *Cybercrime* konsumen, penipuan internet, kejahatan *online*, dan *e-crime*, hal tersebut menyebabkan kesulitan yang cukup serius kepada setiap orang yang terpengaruhi, dan itu bahkan juga bisa berujung pada permasalahan keuangan, yang mana beberapa korban sudah menemui hal tersebut.

Beberapa bentuk penipuan *online* yang dapat kita temukan beberapa diantaranya adalah (Bennett, 2015):

1. *Phising* merupakan penipuan yang bersifat mencuri informasi individu dan terjadi lebih sering pada *email*, dimana korban diminta untuk memverifikasi idTabel untuk, kartu kredit, tanggal lahir, dan *password* yang mana informasi ini sangat berharga.
2. Penipuan *Retail* jenis penipuan yang menjanjikan untuk mendapatkan harga produsen atau pembayaran lebih besar dari harga penjualan yang melibatkan barang mahal seperti mobil, penipu yang menawarkan tawaran ini biasanya datang dari *online* atau melalui internet
3. Penipuan peluang karir merupakan penipuan yang mencari pekerja yang serius secara *online* dan menghubungi serta menawarkan pekerjaan mengurus pembayaran pelanggan, penipu biasanya akan meminta data akun bank untuk dijadikan perwakilan keuangan sebagai alasan,

- biasanya penipu akan menargetkan karyawan diluar wilayah, dengan akun bank yang didapatkan si penipu, penipu akan mencuri uang dan informasi akun bank dan menggunakan informasi pribadi yang dapat diterima melalui akun bank,dan hal yang lebih buruk dapat terjadi .
4. Penipuan menghasilkan uang di rumah dengan komputer – penipuan jenis ini, tidak semua adalah palsu beberapa diantaranya adalah asli, dalam beberapa kasus penipuan ini biasanya akan memberikan program yang akan digunakan untuk menghasilkan uang, akan tetapi program tersebut berisi iklan yang mana bukan menghasilkan uang untuk si pengguna akan tetapi si penipu yang akan mendapatkan penghasilan.
 5. Penipuan *ransomware* merupakan penipuan dimana komputer akan diinfeksi oleh virus *ransomware* dimana korban akan diharuskan untuk membayar biaya untuk membuka data mereka yang penting, dalam kasus ini pembayaran malah hanya akan memperburuk, data korban hilang dan uang yang dibayarkan tidak akan kembali.

Terdapat banyak sekali jenis penipuan *online*, salah satu penipuan *online* dengan total kerugian terbesar adalah penipuan *online* pada *e-commerce* dengan penipuan jenis *Account take-over* pada transaksi *online* yang termasuk pada penipuan *e-commerce* secara *phising* dan pencurian data (*Identity Theft*).

Dari beberapa penipuan *online*, terutama *Account take-over* menargetkan akun dan pencurian kartu kredit yang berujung pada penipuan kartu kredit (Steele, 2017). Penipuan kartu kredit merupakan salah satu jenis pencurian atau penipuan yang melibatkan sebuah kartu kredit, tujuan dari penipuan kartu kredit adalah untuk membeli barang atau jasa tanpa membayar, atau untuk mencuri uang dari akun kartu kredit orang lain (Thomas, 2018).

Jenis – jenis penipuan kartu kredit menurut Michael (Bennett, 2015) adalah sebagai berikut:

1. *Application Fraud* merupakan jenis penipuan yang secara umum terjadi bersama dalam *identity theft*, terjadi ketika orang lain membuat kartu kredit atas nama korban, orang tersebut biasanya pertama kali akan mencuri dokumen pendukung, yang mana kemudian digunakan untuk menggantikan penipuan aplikasi.
2. *Electronic or Manual Credit Card Imprints* merupakan bentuk berikutnya dari penipuan kartu kredit dialami melalui pengandaan kartu kredit, yang mana artinya seseorang melakukan

- pembacaan informasi yang diletakkan pada strip kartu kredit, kemudian data ini digunakan untuk *encode* kartu palsu atau untuk menyelesaikan penipuan transaksi.
3. *CNP (Card Not Present) Fraud* merupakan jenis penipuan kartu kredit yang mana orang tersebut menggunakan kartu korban tanpa secara fisik memiliki kartu tersebut, biasanya pencurian terjadi apabila pelaku tersebut mengetahui nomor akun kartu kredit korban dan menggunakan kartu kredit untuk transaksi di pelaku.
 4. *Lost and Stolen Card Fraud* merupakan jenis penipuan dimana kartu korban hilang dan , kartu tersebut apabila ditransaksikan secara mesin pembayaran biasanya lebih sulit, akan tetapi pembayaran dapat dilakukan apabila digunakan secara *online*, kemudian akan digunakan untuk melakukan pembayaran oleh pelaku untuk transaksi pelaku itu sendiri.
 5. *Fake Card* merupakan jenis penipuan dimana pelaku membuat sebuah kartu kredit yang palsu yang tidak terhubung dengan user manapun dan uniknya kartu tersebut dapat digunakan untuk bertransaksi *online*, perusahaan kartu kredit biasanya tidak akan membayar transaksi yang tidak terhubung dengan profil penggunanya yang spesifik, kasus ini apabila sudah disadari oleh *merchant* biasanya pelaku sudah hilang dengan barang yang sudah dibelinya dengan kartu kredit palsu tersebut.
 6. *Account Takeover* merupakan bentuk yang paling umum dari penipuan kartu kredit, secara dasar, seorang criminal entah bagaimana akan mampu mendapatkan semua informasi korban dan dokumen yang relevan, hal ini biasanya dilakukan secara *online*, pelaku kemudian akan menghubungi perusahaan kartu kredit kemudian berpura – pura untuk menjadi korban, menanyakan perusahaan untuk mengganti alamat, kemudian akan diverifikasi kembali datanya untuk pencocokan, akan tetapi pelaku sudah mendapatkan semua informasi korban sehingga verifikasi tidak akan menjadi masalah besar bagi pelaku, kemudian kartu kredit baru akan dikirim ke alamat yang palsu, kemudian pelaku bisa membuat pembayaran serta transaksi yang tidak diketahui oleh pemiliknya sendiri.

2.4 *Machine Learning*

Machine Learning (ML) adalah ilmu dari komputer untuk belajar dan beraksi seperti manusia, dan mengimprovisasi kemampuan belajar seiring berjalannya waktu dalam aplikasi *autonomous*, dengan memberikan data dan informasi dalam bentuk penelitian dan interaksi dunia nyata (Faggella, 2019).

Menurut beberapa ahli *Machine Learning*, definisi dari *Machine Learning* adalah:

1. *Machine Learning* secara dasar merupakan latihan dari penggunaan algoritma untuk *parse* data, dan belajar dari informasi tersebut, kemudian membuat sebuah determinasi atau prediksi tentang sesuatu di dunia (Copeland, 2016).
2. *Machine Learning* berdasarkan algoritma yang dapat belajar dari data tanpa bergantung pada pemrograman yang berbasis peraturan (Quarterly, 2015).
3. *Machine Learning* merupakan algoritma yang mampu mengetahui cara mengeksekusi tugas yang penting dari generalisasi contoh-contoh yang ada (Mitchell, 2006).

Tipe – tipe pada *Machine Learning* terdiri dari empat tipe yaitu (GN, 2018):

1. *Supervised Learning* yaitu tipe *Machine Learning* yang mampu menempatkan nilai *input* terhadap *output*, dan menghasilkan nilai hasil yang sesuai dengan model yang ditraining terhadap *Machine Learning*, secara khas digunakan untuk regresi dan klasifikasi, dan permodelan prediksi.
2. *Unsupervised Learning* yaitu tipe *Machine Learning* yang mampu menghasilkan keluaran sesuai model tanpa harus ditraining, secara khas digunakan untuk algoritma *Clustering*, algoritma *learning*, *Anomaly Detection* dan permodelan deskriptif.
3. *Semi-supervised Learning* yaitu tipe *Machine Learning* diantara *supervised* dan *unsupervised learning*, dimana kombinasi diantara kedua tipe diterapkan untuk mendapatkan hasil yang diinginkan.
4. *Reinforced Learning* yaitu tipe *Machine Learning* yang berada pada lingkungan training dengan metode *trial* dan *error*, dilatih untuk membuat keputusan yang lebih spesifik, *Machine Learning* tersebut belajar dari pengalaman terdahulu dan mencoba mendapatkan kemungkinan pengetahuan terbaik dan membuat keputusan akurat berdasarkan *feedback* yang diterima.

Ada banyak tipe *Machine Learning* yang berbeda, dengan ratusan tipe *machine learning* dipublikasikan, secara khas *Machine Learning* dikelompokan berdasarkan cara pembelajaran seperti *supervised learning*, *unsupervised learning*, *semi-supervised learning*, *reinforced learning* atau dengan persamaan dalam bentuk atau fungsi seperti klasifikasi, regresi, *Decision Tree* dan sebagainya, diluar dari cara pembelajaran atau fungsi setiap kombinasi dari algoritma machine learning terdiri dari bagian berikut:

1. *Representation* yang merupakan kumpulan klasifikasi atau bahasa yang dimengerti oleh komputer.
2. *Evaluation* yang merupakan fungsi objektif atau penilaian.
3. *Optimization* yang merupakan pemanfaatan algoritma dengan klasifikasi yang diterapkan dengan menggunakan metode optimasi khusus.

Tabel 2.1 The Three Components of Learning Algorithms (Faggella, 2019)

Representation	Evaluation	Optimization
K-nearest neighbor	Accuracy / Error rate	Combinatorial optimization
Support vector machines	Precision and recall	Greedy search
Hyperplanes	Squared error	Branch-and-bound
Naïve Bayes	Likelihood	Continuous optimization
Logistic Regression	Posterior probability	Unconstrained
Decision trees	Information gain	Gradient descent
Set of rules	K-L divergence	Conjugate gradient
Propositional rules	Cost / Utility	Quasi-newton methods
Logic programs	Margin	Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian Network		
Conditional random fields		

Limitasi yang ada pada *Machine Learning* terbagi menjadi dua yaitu:

1. *Overfitting* merupakan limitasi dimana model pada *Machine Learning* mempelajari detil dan *noise* pada training data terlalu jauh hingga mempengaruhi kinerja dari model pada data baru (Brownlee, 2016).
2. Dimensionality merupakan limitasi dimana algoritma machine learning dengan fitur yang semakin banyak akan menyebabkan bekerjanya algoritma pada dimensi yang semakin besar , dan menyebabkan pemahaman data akan semakin sulit.

Machine Learning membutuhkan klasifikasi untuk berfungsi dimana klasifikasi merupakan sebuah teknik untuk mendeterminasikan suatu class yang mana suatu klas bergantung pada satu atau lebih variabel independen, class sering disebut label, target atau kategori.

Proses klasifikasi secara bersamaan dibagi menjadi dua fase yaitu pelatihan dan pengujian. Pada fase pelatihan, algoritma memiliki akses ke nilai atribut prediktor dan atribut tujuan untuk semua contoh set pelatihan dimana informasi tersebut digunakan untuk membangun *model* klasifikasi. *Model* ini mewakili pengetahuan klasifikasi seperti hubungan antara nilai atribut prediktor dan kelas yang menghasilkan prediksi kelas dari contoh yang diberikan nilai atribut prediktornya. Pada fase pengujian, set pengujian nilai-nilai kelas dari contoh tidak ditampilkan. Pada fase pengujian, algoritma diizinkan untuk melihat kelas aktual dari contoh yang baru diklasifikasikan setelah prediksi selesai dibuat. Salah satu tujuan utama dari algoritma klasifikasi adalah untuk memaksimalkan akurasi prediksi yang diperoleh oleh *model* klasifikasi ketika mengklasifikasikan contoh dalam set pengujian yang tidak terlihat selama pelatihan. Algoritma klasifikasi terdiri dari *Decision Tree*, *K-Nearest Neighbor*, *Support Vector Machines*, *Bayesian Network*, *Neural Network* dan lain-lain (Neelamegam & Ramaraj, 2013).

2.5 Confusion Matrix

Confusion Matrix adalah alat visualisasi yang biasa digunakan pada *supervised learning* dengan hasil seperti Tabel 2.2 di bawah ini (Gorunescu, 2011).

Tabel 2.2 Model Confusion Matrix (Han, 2012)

Correct Classification	Classified as	
	Positives	Negatives
Positives	True Positives	False Negatives
Negatives	False Positives	True Negatives

Hasil dari *Model Confusion Matrix* sebagai berikut:

1. TP (*True Positive*) adalah jumlah *record* positif yang diklasifikasikan sebagai positif. Contohnya suatu transaksi aslinya *fraud* positif dan diklasifikasi hasilnya *fraud* positif termasuk TP (Han, 2012).

2. FP (*False Positive*) adalah jumlah *record* negatif yang diklasifikasikan sebagai positif. Contohnya suatu transaksi aslinya *non-fraud* negatif tetapi diklasifikasi hasilnya *fraud* positif termasuk FP (Han, 2012).
3. FN (*False Negative*) adalah jumlah *record* positif yang diklasifikasikan sebagai negatif. Contohnya suatu transaksi aslinya *fraud* positif tetapi diklasifikasi hasilnya *non-fraud* negatif termasuk FN (Han, 2012).
4. TN (*True Negative*) adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif. Contohnya suatu transaksi aslinya *non-fraud* negatif dan diklasifikasi hasilnya *non-fraud* negatif termasuk TN (Han, 2012)

Berikut kriteria untuk mengevaluasi tingkat efektifitas pada *binary* klasifikasi :

1. *Precision*

Precision (P) adalah jumlah elemen data dengan klasifikasi yang benar dan termasuk *fraud* positif dibagi dengan total elemen data yang termasuk *fraud* positif (Ling, et al., 2014)..

Perhitungan *Precision* (Sokolova & Lapalme, 2009)

$$Precision = \frac{TP}{(TP+FP)} \quad (1)$$

2. *Recall*

Recall (R) adalah jumlah elemen data dengan klasifikasi yang benar dan termasuk *fraud* positif dibagi dengan total elemen data *fraud* positif yang sebenarnya (Ling, et al., 2014).

Perhitungan *Recall* (Sokolova & Lapalme, 2009)

$$Recall = \frac{TP}{(TP+FN)} \quad (2)$$

3. *Accuracy*

Accuracy adalah efektivitas keseluruhan suatu *classifier* dalam bentuk persen (Sokolova & Lapalme, 2009).

Perhitungan *Accuracy* (Sokolova & Lapalme, 2009)

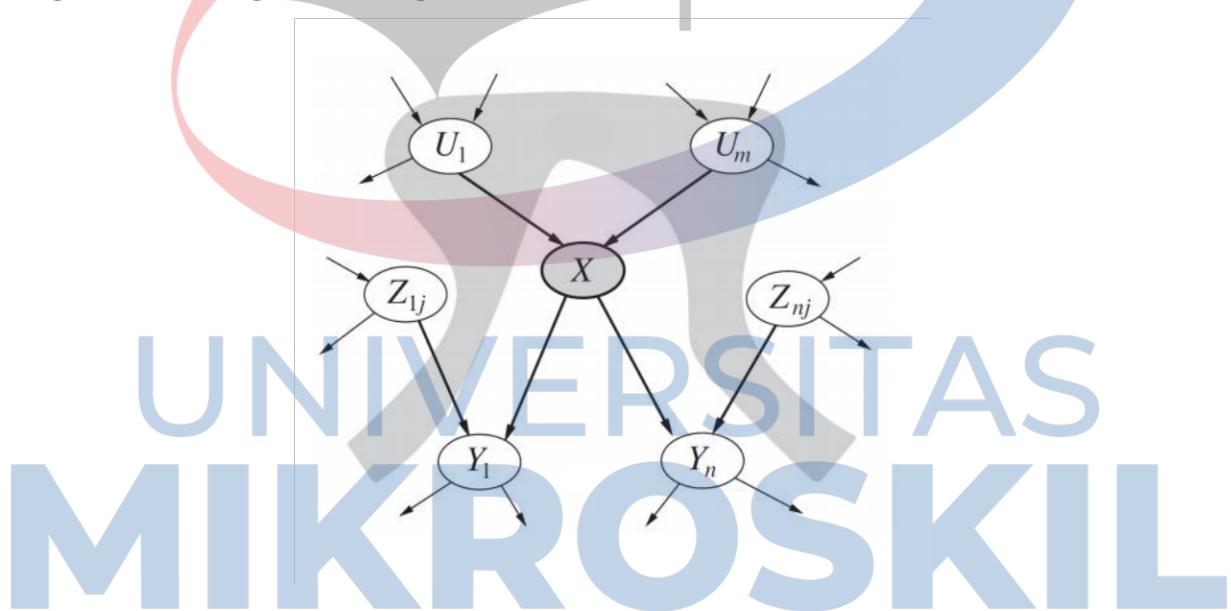
$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (3)$$

2.6 Bayesian Network

Bayesian Network merupakan sebuah tipe probabilistik bermodelkan *Directed Acyclic Graph* (DAG) yang menggunakan *Bayesian Inference* untuk komputasi probabilitas, *Bayesian Network* bertujuan untuk memodelkan kondisi yang saling bergantung, dan sebab akibat dengan representasi ketergantungan kondisi oleh bagian *node* dalam grafik yang terarah, melalui hubungan ini, dapat secara efisien melakukan inferensi pada variable acak dalam grafik melalui faktor penggunaan (Soni, 2018) dengan rumus JPD (*Joint Probability Distribution*) dinyatakan sebagai persamaan berikut:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i)) \quad (4)$$

Menggunakan hubungan yang dispesifikasikan oleh *Bayesian Network*, dapat mengambil sebuah representasi yang terfaktorisasi dan padat dari penggabungan distribusi probabilitas dengan mengambil keuntungan dari independansi berkondisi.



Gambar 2.2 *The Bayesian Network* (Soni, 2018)

Bayesian Inference merupakan seperangkat alat yang sangat kuat untuk memodelkan variabel *random*, seperti nilai dari parameter regresi, statistik demografik, bagian kata dari kalimat (Kramer, 2016).

Bayesian Infererence terdapat dalam dua bentuk, pertama dengan sederhana mengevaluasikan gabungan probabilitas dari nilai yang ditentukan dari setiap variable atau subset dalam *network*, evaluasi produk menggunakan probabilitas kondisi yang disediakan, kedua yaitu dengan tugas *inference* yang lebih menarik yaitu menemukan $P(x|e)$, atau mencari probabilitas dari beberapa tugas dari variabel subset (x) diberikan tugas dari variable lain (e) (Soni, 2018).

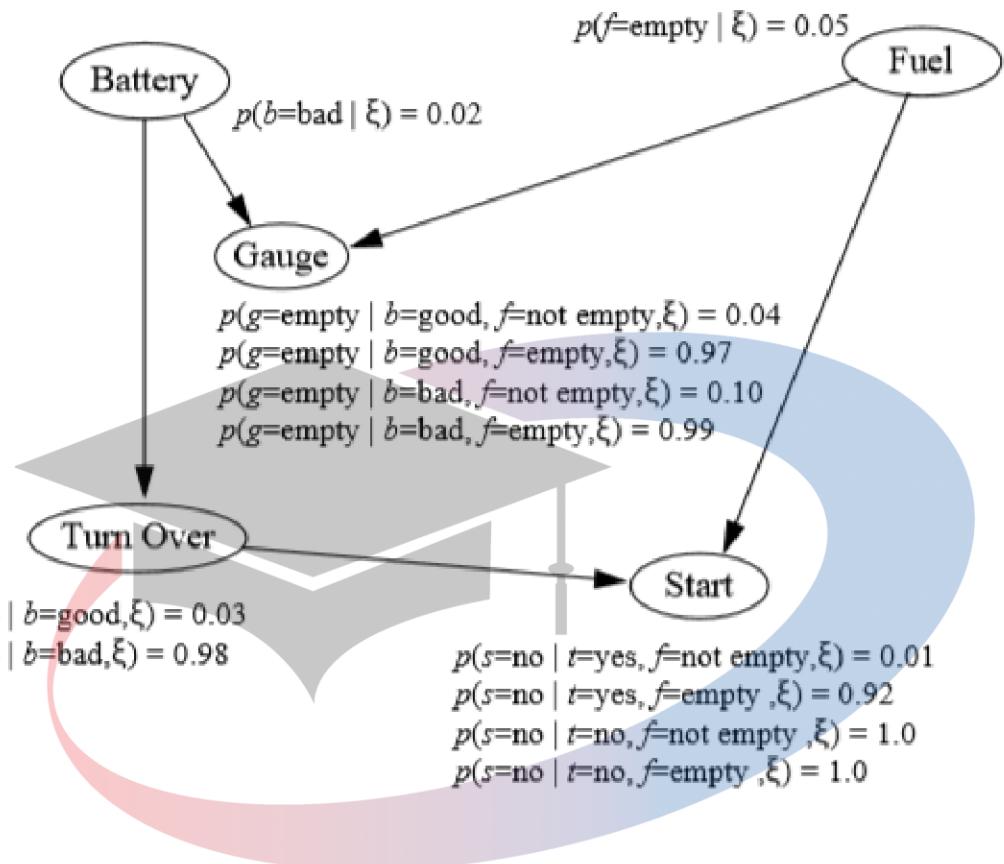
Cara kerja *Bayesian Network* ialah dengan *machine learning* menggunakan tipe *supervised learning* dimana sebelum menentukan probabilitas digunakan dataset training untuk algoritma untuk hasil sesuai dengan model yang dilatih, penentuan prediksi dilakukan dengan cara penentuan domain dan variabel acaknya sesuai dengan jumlah *event* sesuai dengan persamaan 2.

$$U = \{x_1, \dots, x_n\} \quad (5)$$

Simbol U untuk *bayesian network* merepresentasikan JPD kepada setiap *node* yang berada pada X_1 hingga X_n (Heckerman, 2008), apabila penentuan variabel acak sudah dilakukan maka tahap berikutnya ialah tentukan *event* pada setiap variabel acak, dimana setiap variabel acak harus memiliki hubungan yang berlogika dan merepresentasikan suatu *event*, setelah menentukan variabel acak, ditentukan tipe variabel pada variabel acak, dimana variabel dengan nilai diskrit, yaitu variabel dengan jumlah nilai tertentu dan terbatas (Stephanie, 2013) akan dikategorikan sebagai variabel diskrit dan variabel dengan nilai berlanjut, yaitu variabel yang memiliki nilai yang berubah – ubah dan dapat bervariasi serta memiliki nilai yang tidak terbatas (Stephanie, 2013), akan dikategorikan sebagai variabel berlanjut dan setelah menentukan tipe variabel pada variabel acak maka selanjutnya dibuat sebuah model berdasarkan *event* seperti pada gambar 2.6, dan juga konstruksi CPT, kemudian untuk melakukan komputasi probabilitas terhadap gambar 2.6 yaitu pada variabel f apabila diberikan c adalah tidak yang kemudian dinyatakan sebagai persamaan berikut (Heckerman, 2008):

$$P(f|s = no, \epsilon) = \frac{P(f, s = no | \epsilon)}{P(s = no | \epsilon)} = \frac{\sum_{b,g,t} P(b,f,g,t,s=no | \epsilon)}{\sum_{b,f,g,t} P(b,f,g,t,s=no | \epsilon)} \quad (6)$$

Model dari Heckerman (Heckerman, 2008) dimana penentuan variabel acak, event dan juga CPT yang dibuat digambarkan pada gambar berikut:



Gambar 2.3 Contoh Model Bayesian Network (Heckerman, 2008)

Aturan pembuatan model, dimana setiap *node* yang berhubungan harus memiliki hubungan berlogika dan merepresentasikan suatu *event*, pada gambar 2.3 apabila *Battery* adalah B dan *Turn Over* adalah T dan *Start* adalah S, secara hukum *Bayesian Network*, B menyebabkan T dimana S disebabkan T dan hukum ini berlaku untuk setiap *node* yang ada pada *Bayesian Network*, kemudian untuk setiap variabel yang sudah ditentukan tipe nodenya dimana node terbagi menjadi 2 yaitu node dengan *parent* dan *node* tanpa *parent* yang kemudian akan dihitung, dengan cara berikut:

1. Setiap *node* yang tidak memiliki parent seperti pada gambar 2.3 dimana Variabel B dan F merupakan *input* probabilitas dimana *input* probabilitas berdasarkan suatu data.
2. Untuk *node* yang memiliki parent akan dihitung JPD (*Joint Probability Distribution*) dimana JPD dihitung menggunakan CPT (*Conditional Probability Table*).
3. Untuk setiap *node* dengan parent yang sudah terhitung JPD nya dan pada *node* terbawah yaitu *main event* akan dihitung probabilitas *fraud* yang merupakan gabungan dari beberapa JPD dari *parent* teratas sesuai dengan persamaan 3.

Pada proses pengujian *Bayesian Network* akan digunakan data transaksi, pembayaran dan pelaporan pengguna yang akan dijadikan sebagai *input* probabilitas berdasarkan data dan CPT pengujian akan ditentukan sebelumnya untuk menghasilkan JPD setiap *node* dan akan digabungkan untuk menghitung probabilitas *fraud*. Pada proses pengujian akurasi *Bayesian Network* akan digunakan dataset dari sampel dengan 25 elemen data dimana salah satu elemen yaitu *default_payment_next_month* akan digantikan dengan nilai dari perhitungan *Bayesian Network* dan hasil akan dibandingkan dengan data asli dimana hasil akurasi akan dinyatakan dalam bentuk persentase dengan batas nilai 0 sampai 100 persen sebagai berikut.

UNIVERSITAS MIKROSKIL