

BAB II

LANDASAN TEORI

2.1 Pengenalan Kepribadian Menggunakan DISC

2.1.1 Sejarah DISC

Menurut (Mursalin, A. 2013). Pada awal abad 19, para ahli psikologi dan ilmu-ilmu psikologi hanya mempelajari dan membahas perilaku serta problema orang-orang yang menyimpang saja (gangguan mental). Pada saat itu belum ada perhatian terhadap sikap dan perilaku orang normal. Barulah pada tahun 1928 seorang konsultan psikologi dan dosen yaitu Dr. William Moulton Marston membuat buku yang berjudul “*The Emotions of Normal People*”. Beliau tampaknya lebih dahulu mulai menyadari bahwa orang normal-pun ternyata memiliki sisi-sisi yang menarik karena perbedaan *personality*-nya. Pada tahun 1921, C.G. Jung mulai memperkenalkan empat tipe yang berorientasi pada fungsi psikologi yaitu *thinking* (berpikir), *feeling* (perasaan), *sensation* (sensasi) dan *intuition* (intuisi). Terpengaruhi oleh kebiasaan empat unsur tersebut atau memang sengaja mempermudah Marston membuat empat tipe dasar manusia yang dilambangkan dengan empat faktor penting dalam mengukur kepribadian. Faktor penting tersebut adalah *Dominant* (sifat dominan/kuasa), *Influence* (sifat mempengaruhi), *Steadiness* (sifat stabil) dan *Compliance* (sifat pemikir/kritis) atau yang biasa disingkat *DISC*.

Dengan dimulainya era ilmiah maka Dr. Marton membuat kuesioner yang digunakan sebagai alat untuk mengidentifikasi tipe orang, apakah seseorang termasuk tipe *D*, *I*, *S* atau *C*. Pada masa kini, psikologi *modern* dan bahkan dengan menggunakan komputer analisis atas kuesioner tersebut dapat dihasilkan data identifikasi dengan lebih cepat. Seperti umumnya alat-alat tes sejenis (termasuk *IQ-test*), *DISC* pertama kali digunakan untuk kepentingan militer dan secara luas digunakan sebagai bagian dalam proses penerimaan tentara AS pada tahun-tahun

menjelang Perang Dunia II. Setelah keandalannya terbukti, kemudian DISC secara bertahap dipakai untuk kepentingan rekrutmen yang lebih umum. Teori *DISC* merupakan teori yang saat ini paling banyak dipakai dan dibahas di Indonesia, kemungkinan hal ini dikarenakan kemudahan dalam mengukur dan menentukan kepribadian dengan teori ini.

2.1.2 Macam-Macam Tipe *DISC*

Berdasarkan teori *DISC*, kepribadian seseorang digolongkan menjadi empat tipe yaitu (Marston, 1928) :

1. *Dominance* (Dominan)

Orang dengan sifat *dominance* (kuasa), suka mencapai hasil dengan cepat, suka akan tantangan dan hal bidang baru serta mengambil otoritas. Mereka akan merasa nyaman pada lingkungan penuh kekuasaan dan wewenang dengan kesempatan untuk meningkatkan prestasi individu. Orang dominan suka bertanggung jawab dan tidak ingin berada di bawah kendali orang lain dan paling nyaman ketika mereka dapat mengontrol lingkungan. Mereka bekerja dengan baik ketika mereka bebas dari kontrol dan pengawasan. Orang dominan selalu memunculkan ide besar, sehingga sering memegang kendali dan membuat keputusan yang cepat. Akan tetapi orang dominan sering melewatkan detail, kurang berkomitmen, dan sering tidak memperdulikan penilaian dan perasaan orang lain. Sehingga orang dengan *dominance* kuat, membutuhkan orang lain sebagai penasihat, menghitung resiko dan fakta penelitian. Biasanya orang dengan kecenderungan dominan akan mampu mengemban tugas sebagai pimpinan, misalnya CEO atau Presiden.

2. *Influence* (Mempengaruhi)

Individu dengan pengaruh *influence* perilakunya cenderung cepat, akan tetapi lebih suka berada dan bekerja sama dengan orang lain ketika mengerjakan sebuah tugas. Mereka suka menghibur, berpengaruh dan

pintar mengambil hati orang. Menjadi terkenal dan dielu-elukan banyak orang adalah hal yang mereka sukai. Mereka sangat senang bergaul dengan komunitas atau suatu kelompok untuk memperbanyak relasi. Orang *influence* menikmati berhubungan dengan orang lain dan membuat kesan baik, berbicara lantang dan menciptakan lingkungan yang positif dan antusias. Orang dengan sifat *influence* lebih suka dalam bidang pembinaan dan konseling. Sayangnya, orang dengan tipe ini memiliki perasaan yang halus dan tidak berkonsentrasi dengan tugas yang sekarang dihadapi. Mereka lebih cocok sebagai penasehat. Mereka membutuhkan orang lain untuk mencari fakta, berkomunikasi lengkap, menghormati ketulusan, memberi penghargaan atas hal-hal kecil, melakukan pendekatan secara logis.

3. *Steadiness* (Stabil)

Seseorang dengan *steadiness* dikenal dengan konsistensinya melakukan sesuatu sampai berhasil/selesai, tidak berorientasi pada kecepatan tetapi konsistensi. Paling baik dalam menciptakan harmoni di tempat kerja. Kontribusi paling positif untuk orang *steadiness* ini adalah menjadi pendengar sejati, sabar, suka membantu dan pandai mengendalikan keadaan. Individu dengan kategori ini selalu fokus pada bekerja sama dengan orang lain untuk menyelesaikan tugas mereka. Kelemahannya adalah tidak memiliki motivasi diri yang kuat, sehingga cepat berubah dan mudah dipengaruhi orang lain dan enggan membuat keputusan. Tipe ini memerlukan orang yang dapat menekan mereka dan membantu memprioritaskan tugas, pekerjaan dan memiliki fleksibilitas tinggi dalam prosedur kerjanya. Dalam rangka memiliki efektifitas yang optimal, individu ini perlu diberitahu secara mendalam tentang perubahan yang akan datang sesegera mungkin agar dapat menyesuaikan. Orang *steady* akan nyaman jika berada pada lingkungan yang minimal konflik, berorientasi pada tugas

kelompok, memiliki penghargaan yang tulus dan percaya pada kemampuan orang lain.

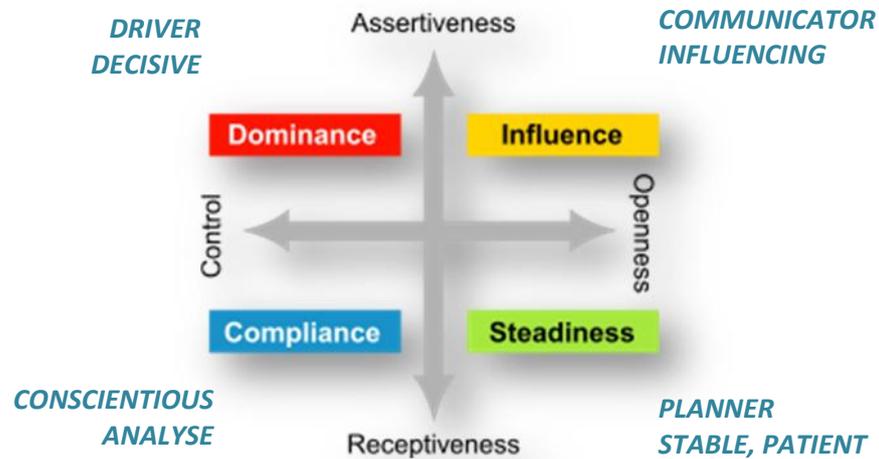
4. *Compliance* (Pemikir)

Compliance menghasilkan pekerja tekun, fokus pada kualitas dan ketepatan. Ciri orang dengan sifat kuat *compliance* adalah sangat teliti dan suka berpikir rumit. Tipe ini akan nyaman jika berada dalam lingkungan yang memiliki *ekspektasi* kinerja yang jelas. Mereka senang melakukan apa saja yang berdampak baik. Orang *compliance* biasanya teguh dalam pendirian dan pilihannya. Terkadang terkesan ngotot. Sifat positif yang dapat diambil dari *compliance* adalah teliti, berfikir kritis, menggunakan pendekatan secara halus dan analitis, memiliki rencana yang matang, dapat menyelesaikan masalah dengan baik, *profesional*, diplomatis dan punya loyalitas tinggi. Kelemahannya adalah cenderung meragu dan terkesan lambat dalam pengambilan keputusan karena terlalu teliti. Maka, pekerjaan yang dilakukan akan terproses dengan lambat, pendendam dan terlalu kritis. Orang yang sangat teliti, akan membutuhkan orang lain untuk berkompromi dan mengambil keputusan yang cepat. Sehingga orang *compliance* biasanya bekerja di bawah komando orang *dominance* sebagai Wakil Presiden atau Asisten Manajer.

2.1.3 Konsep Teori DISC

Kepribadian berdasarkan profil DISC adalah perilaku yang didasarkan dari *stimulus* dan respon dimana setiap orang akan berperilaku berbeda terhadap respon yang didapat. Perilaku adalah jumlah dari gabungan berbagai respon yang dilakukan dari beragam *stimulus* yang diberikan. Dasar teori DISC adalah biasa disebut ***Biaxial Model*** atau model dua sumbu (*axis*) perilaku, merupakan pergerakan satu sumbu dengan sumbu lain mengenai perilaku yang saling berlawanan. Dua sumbu tersebut merupakan sifat yang saling berlawanan. Dua sumbu yang saling

berlawanan pada tes DISC ini digunakan yaitu antara *Assertiveness* dan *Receptiveness* dan antara *Openness* dan *Control*.



Gambar 2.1 Biaxial Model DISC

I. Model Sumbu Perilaku *Assertive - Receptiveness*

1. Assertiveness

Sifat orang yang terbuka (proaktif) dan dapat langsung mengemukakan keterbukaan diri. Berkecenderungan untuk memimpin daripada mengikuti dan serta merta dapat bertindak apabila memungkinkan. Tipe asertif akan cepat merasa mendapat kesempatan yang tidak boleh disia-siakan, mandiri dalam bertindak, cenderung memberi perintah daripada melakukan sendiri. Lebih memberikan instruksi dibandingkan kerjasama.

2. Receptiveness

Sifat yang menunjukkan kesabaran dan kehati-hatian, cenderung menghindari resiko dan jarang bertindak tergesa-gesa apabila masih memungkinkan. Menghindari perubahan dan kejutan, lebih memilih ketenangan dan situasi yang dapat diperkirakan.

II. Model Sumbu Perilaku *Openness - Control*

1. *Openness*

Sifat model DISC untuk mengukur sikap sosial, pendekatan yang berbeda terhadap interaksi dengan orang lain. *Openness* adalah orang yang sangat terbuka, ramah, bersahabat. Mudah mengekspresikan diri dan memiliki hubungan yang sangat kuat dengan orang lain. Orang yang terbuka dapat bekerja pada situasi kerja yang sangat emosional dengan mengekspresikan perasaan.

2. *Control*

Control adalah sifat sebaliknya dimana menggunakan dasar rasio di atas perasaannya. Pendekatan kepada orang lain lebih menggunakan ide sendiri dibandingkan dengan gaya dari orang lain. Sifat ditunjukkan dengan ketidakpercayaan, curiga, tidak mudah menunjukkan perasaan kepada orang lain.

III. Biaxial Model DISC

Dua kutub yang saling berbeda adalah dasar dari profil DISC yang dikembangkan. Pada tiap sumbu yang saling berdekatan menunjukkan karakteristik dari setiap elemen dari tes DISC. Pada aspek *Dominance* seperti contoh merupakan pendekatan antara sifat *Assertiveness* dan *Control*. Pada setiap sumbu antara *Assertiveness-Receptiveness* dan *Openness-Control* terdapat area-area yang dapat didefinisikan secara lebih jelas melalui berbagai karakteristik yang dapat didefinisikan sebagai berikut:

- a. ***Driver***, area paling kiri-atas antara perilaku *Assertiveness* dan *Control*. Menunjukkan faktor *Dominance* dari konsep DISC yang diterangkan sebagai

- pengarah dengan menunjukkan karakteristik orang yang bermotivasi tinggi untuk berhasil dengan sifat-sifat kompetitif terhadap orang lain.
- b. **Communicator**, area paling kanan-atas kombinasi dari perilaku *Assertiveness* dan *Openness* dan pada konsep DISC diistilahkan *Influence*. Tipe orang seperti ini adalah komunikatif dan bersifat sosial. Akrab dan terbuka dengan orang lain dan mudah menyesuaikan diri dengan lingkungan baru.
 - c. **Planner**, area paling kanan-bawah digambarkan sebagai *steadiness* dengan karakteristik khusus adalah kesabaran, bertahan pada sikap atau nilai, tidak menyukai perubahan drastis, membutuhkan waktu untuk merencanakan dengan kehati-hatian.
 - d. **Analyst**, area paling kiri-bawah diistilahkan sebagai *compliance*. *Analyst* diterangkan sebagai berwawasan, berupaya mengetahui. Gabungan antara terkontrol dan kesiapan menunjukkan pribadi yang teratur, sebisa mungkin mengikuti aturan. Kecenderungan pada ketepatan dan kejelasan.

2.1.4 Metode Pengukuran DISC

Banyak metode untuk melakukan pengukuran DISC. Umumnya sudah tersedia aplikasi dengan berbasis *online*. Namun cara pengukuran DISC yang dimiliki oleh PT Indonesia Spring menggunakan metode sendiri sesuai dengan standar sistem perusahaan, yaitu dengan tes tabel perilaku (tabel tes tipologi), tes kuesioner dan wawancara psikotes DISC oleh tenaga ahli dari HRD. Metode ini telah lama diterapkan di perusahaan dan menjadi *assessment tool* yang efektif namun membutuhkan waktu dan proses cukup panjang dalam interpretasi hasilnya.

Mengenai metode pengukuran DISC ini telah terangkum pada buku modul “Basic Management Program” perusahaan. Contoh pengambilan angka tabel tes tipologi dapat dilihat pada Gambar 2.2 sebagai berikut,

• Untuk perilaku yang PALING SESUAI dengan anda	→ beri nilai “1”
• Untuk perilaku paling SESUAI TINGKAT BERIKUTNYA	→ beri nilai “2”
• Untuk tingkat LEBIH RENDAH	→ beri nilai “3”
• Untuk perilaku YANG PALING SEDIKIT KESESUAIANNYA dengan anda	→ beri nilai “4”

Contoh :

memarahi **3** memahami **4** mendiamkan **1** memikirkan **2**

	Kolom A	Kolom B	Kolom C	Kolom D	
I.	Mengarahkan 2	Mempengaruhi 1	Bersahabat 3	Hati-hati 4	A =
II.	Pasti 1	Optimistis 2	Sabar 3	Menahan diri 4	B =
III.	Berani 1	Antusias 2	Menenangkan 4	Analitis 3	C =
IV.	Kompetitif 1	Senang bicara 2	Mendengarkan 4	Persis-akurat 3	D =
V.	Kuat-tegas 1	Mempesona 2	Kalem 4	Ingin tahu 3	
Kali 5	30 Jumlah 6	45 Jumlah 9	90 Jumlah 18	85 Jumlah 17	

Gambar 2.2 Tabel Tes Identifikasi Perilaku (Tipologi)

Dijelaskan bahwa Kolom A mewakili model biaxial model dengan tipe karakter *Driver*, Kolom B untuk tipe *Communicator*, Kolom C untuk tipe *Planner* dan Kolom D untuk tipe *Analyst*. Data angka tipologi tersebut nantinya akan didukung oleh data tes kuesioner dan wawancara pada tahapan proses selanjutnya.

2.2 Data Mining

Tan (2006) mendefinisikan *data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga *knowledge discovery*. Selanjutnya perbedaan *data mining* dan *data warehouse* adalah *data mining* merupakan bidang yang sepenuhnya menggunakan apa yang sepenuhnya digunakan oleh *data warehouse*, bersama dengan bidang yang menangani masalah pelaporan dan manajemen data. Sementara *data warehouse* sendiri bertugas untuk menarik/meng-*query* data dari basis data mentah untuk memberikan hasil data

yang nantinya digunakan oleh bidang yang menangani manajemen, pelaporan, dan *data mining*. Dengan *data mining* inilah, penggalian informasi baru dapat dilakukan dengan bekal data mentah yang diberikan oleh *data warehouse*. Hasil yang diberikan oleh ketiga bidang tersebut berguna untuk mendukung aktivitas bisnis cerdas (*business intellegent*).

Pekerjaan yang berkaitan dengan *data mining* dibagi menjadi 4 (empat) kelompok, yaitu (Prasetyo, E. 2012) :

1. *Prediction Modelling* (model prediksi), berkaitan dengan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat.
2. *Cluster Analysis* (analisis kelompok), pengelompokan data-data kedalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada.
3. *Association Analysis* (analisis asosiasi), digunakan untuk menemukan pola yang menggambarkan kekuatan fitur dalam data dengan tujuan untuk menemukan pola yang menarik dengan cara yang efisien.
4. *Anomaly Detection* (deteksi anomali), berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain.

2.2.1 Tugas Utama Data Mining

Secara umum *data mining* mempunyai tugas utama :

I. Klasifikasi (*Classification*)

Klasifikasi bertujuan untuk mengklasifikasikan *item* data menjadi satu dari beberapa kelas standart. Sebagai contoh, suatu program email dapat mengklasifikasikan email yang sah dengan email spam. Beberapa algoritma klasifikasi antara lain *pohon keputusan*, *nearest neighbor*, *naïve bayes*, *neural networks*, dan *support vector machines*.

II. Regresi (*Regression*)

Regresi merupakan permodelan dan investigasi hubungan dua atau lebih variabel. Dalam analisis regresi ada satu lebih variabel independent / prediktor yang biasa diwakili dengan notasi dengan notasi x dan satu variabel respon yang biasa diwakili dengan notasi y (Santosa, 2007).

III. Pengelompokan (*Clustering*)

Clustering merupakan metode pengelompokan sejumlah data ke dalam kluster(*group*) sehingga dalam setiap kluster berisi data yang semirip mungkin.

IV. Pembelajaran Aturan Asosiasi (*Association Rule Learning*)

Pembelajaran aturan asosiasi mencari hubungan antara variabel. Sebagai contoh suatu toko mengumpulkan data kebiasaan pelanggan dalam berbelanja. Dengan menggunakan pembelajaran atur asosiasi, toko tersebut dapat menentukan produk yang sering dibeli bersamaan dan menggunakan informasi ini untuk tujuan pemasaran.

2.3 Klasifikasi (*Classification*)

Salah satu tugas utama *data mining* adalah klasifikasi. Klasifikasi digunakan untuk menempatkan bagian yang tidak diketahui pada data ke dalam kelompok yang sudah diketahui. Klasifikasi menggunakan variabel target dengan nilai nominal. Dalam satu set pelatihan, variabel target sudah diketahui. Dengan pembelajaran dapat ditemukan hubungan antara fitur dengan variabel target. Ada dua langkah dalam proses klasifikasi (Han and Kember, 2006):

1. Pembelajaran (*learning*): pelatihan data dianalisis oleh algoritma klasifikasi.
2. Klasifikasi: data yang diujikan digunakan untuk mengakulasi akurasi dari aturan klasifikasi. Jika akurasi dianggap dapat diterima, aturan dapat diterapkan pada klasifikasi data tuple yang baru.

2.4 Pengukuran Kinerja Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi dengan benar. Namun tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar, sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*), seperti yang dicontohkan dibawah ini :

Tabel 2.1 Matriks Konfusi untuk Klasifikasi Dua Kelas

f_{ij}		Kelas hasil prediksi (j)	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas = 1	f_{11}	f_{10}
	Kelas = 0	f_{01}	f_{00}

Keterangan :

f_{11} = jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1

f_{10} = jumlah data dalam kelas 1 yang secara salah dipetakan ke kelas 0

Berdasarkan matriks konfusi, kita dapat mengetahui jumlah data dari masing-masing kelas yang diprediksi secara benar, yaitu ($f_{11} + f_{00}$), dan data yang diklasifikasi secara salah, yaitu ($f_{10} + f_{01}$). Kuantitas matriks konfusi dapat diringkas menjadi dua nilai, yaitu akurasi dan laju *error*. Dengan mengetahui jumlah data yang diklasifikasi secara benar, kita dapat mengetahui akurasi hasil prediksi, dan dengan mengetahui jumlah data yang diklasifikasi secara salah, kita dapat mengetahui laju *error* dari hasil prediksi yang dilakukan.

Untuk menghitung **akurasi** digunakan formula,

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}} \dots\dots(2.1)$$

Untuk menghitung **laju error** (kesalahan prediksi) digunakan formula,

$$\text{Laju error} = \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{10}+f_{01}}{f_{11}+f_{10}+f_{01}+f_{00}} \dots\dots(2.2)$$

Semua algoritma klasifikasi berusaha membentuk model yang mempunyai akurasi tinggi (laju *error* rendah).

2.5 Teori Bayes

Menurut (Presetyo, E.2012). Bayes merupakan teknik prediksi probalistik sederhana yang berdasar pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidak tergantungan) yang kuat (naif). Dengan kata lain, dalam *Naïve Bayes*, model yang digunakan adalah "model fitur independen"

Dalam Bayes (terutama *Naïve Bayes*), maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidak adanya fitur lain dalam data yang sama. Contohnya, pada kasus klasifikasi hewan dengan fitur penutup kulit, melahirkan, berat dan menyusui. Di sini ada ketergantungan pada fitur menyusui karena hewan yang menyusui biasanya melahirkan, atau hewan bertelur tidak menyusui. Dalam Bayes, hal tersebut tidak dipandang sehingga masing-masing fitur seolah tidak memiliki hubungan apapun.

Prediksi Bayes didasarkan pada teorema Bayes dengan rumus pada 2.3 :

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \dots\dots\dots(2.3)$$

Keterangan :

$P(H | E)$ = Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) E terjadi.

$P(E| H)$ = Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.

$P(H)$ = Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

$P(E)$ = Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis atau bukti yang lain.

Ide dasar dari aturan Bayes adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan pada beberapa bukti (E) yang diamati. Ada beberapa hal penting dari aturan Bayes yaitu :

1. Sebuah probabilitas awal/priori H atau $P(H)$ adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau $P(H|E)$ adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Contoh, dalam suatu prediksi untuk memperkirakan terjadinya kecelakaan, ada faktor yang mempengaruhi terjadinya kecelakaan, yaitu ngebut. Jika diterapkan dalam *Naïve Bayes*, Probabilitas terjadinya kecelakaan, jika bukti ngebut sudah diamati, dinyatakan dengan

$$P(\text{Kecelakaan} | \text{Ngebut}) = \frac{P(\text{Ngebut} | \text{Kecelakaan}) \times P(\text{Kecelakaan})}{P(\text{Ngebut})}$$

$P(\text{Kecelakaan} | \text{Ngebut})$ adalah nilai probabilitas hipotesis kecelakaan terjadi jika bukti ngebut sudah diamati. $P(\text{Ngebut} | \text{Kecelakaan})$ adalah probabilitas bahwa ngebut yang diamati akan mempengaruhi terjadinya kecelakaan. $P(\text{Kecelakaan})$ adalah probabilitas awal kecelakaan tanpa memandang bukti apapun, sementara $P(\text{Ngebut})$ adalah probabilitas terjadinya ngebut.

Teorema Bayes juga bisa menangani beberapa bukti, misalnya ada E_1 , E_2 , dan E_3 sehingga akhir untuk hipotesis (H) dapat dihitung dengan rumus pada 2.4 :

$$P(H | E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3 | H) \times P(H)}{P(E_1, E_2, E_3)} \dots \dots \dots (2.4)$$

Karena yang digunakan untuk bukti adalah independen, bentuk di atas dapat diubah dengan rumus pada 2.5 :

$$P(H | E_1, E_2, E_3) = \frac{P(E_1 | H) \times P(E_2 | H) \times P(E_3 | H) \times P(H)}{P(E_1) \times P(E_2) \times P(E_3)} \dots \dots \dots (2.5)$$

2.6 Metode Naïve Bayes untuk Klasifikasi

(Prasetyo, E. 2012). Kaitan antara Naïve Bayes dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target pertama dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukan dalam

model klasifikasi. Formulasi Naïve Bayes untuk klasifikasi dengan rumus pada 2.6 :

$$P(Y | X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \dots\dots\dots(2.6)$$

Keterangan :

$P(Y | X)$ = probabilitas data dengan vektor X pada kelas Y

$P(Y)$ = probabilitas awal kelas Y

$\prod_{i=1}^q P(X_i|Y)$ = probabilitas independen kelas Y dari semua fitur dalam vektor X

Karena $P(X)$ selalu tetap, sehingga dalam perhitungan prediksi nantinya cukup hanya dengan menghitung $P(Y) \prod_{i=1}^q P(X_i|Y)$ dengan memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara probabilitas independen $\prod_{i=1}^q P(X_i|Y)$ tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y, yang dinotasikan dengan rumus pada 2.7 :

$$P(X|Y=y) = \prod_{i=1}^q P(X_i|Y=y) \dots\dots\dots(2.7)$$

Setiap set fitur $X = \{X_1, X_2, X_3, \dots, X_q\}$ terdiri atas q atribut (q dimensi).

Umumnya, metode *Naïve Bayes* ini mudah dihitung untuk fitur bertipe kategoris seperti pada kasus klasifikasi “jenis kelamin” dengan nilai {pria, wanita}. Namun untuk tipe numerik (kontinu), ada perlakuan khusus sebelum dimasukkan dalam *Naïve Bayes*, yaitu :

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasi fitur kontinu kedalam fitur ordinal.
2. Mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data peralihan. Distribusi Gaussian biasanya dipilih untuk merepresentasikan probabilitas bersyarat dari fitur kontinu pada sebuah kelas $P(X_i|Y)$, sedangkan distribusi Gausaian dikarakteristikan dengan dua parameter. mean μ , dan

varian, σ^2 . Untuk setiap kelas y_j , probabilitas bersyarat kelas y_j untuk fitur X_i dengan rumus pada 2.6 :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp \left[-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right] \dots\dots\dots(2.8)$$

Keterangan :

μ_{ij} : mean sampel X_i (\bar{X}) dari semua data latih yang menjadi milik kelas y_j .

σ_{ij}^2 : varian sampel (S^2) dari data latih, ($S = Stdev$).

2.7 Penelitian Sebelumnya

Naïve Bayes merupakan metode populer yang banyak di gunakan untuk klasifikasi. Berbagai riset yang telah dilakukan berkaitan dengan kasus prediksi yang menggunakan metode *Naïve Bayes*, antara lain :

1. Nur Indah Sari, lulusan Universitas Muhammadiyah Gresik Tahun 2013. Penelitiannya berjudul :“*Klasifikasi Kecendrungan Penyelesaian Studi Mahasiswa Baru dengan Menggunakan Metode Naïve Bayes*”. Atribut yang dilakukan sebagai data latih sistem adalah NP (Nomor Pegawai), Nama, Umur, Jenis Kelamin, Pendidikan, Agama, Level, Lama Kerja , Kesehatan dan Jadwal. Atribut jadwal digunakan sebagai kelas data. Adapun data yang diambil dalam penilitian ini adalah sampel dari 40 *record* dengan kelas “Lama” dan “Tidak Lama” masing-masing berjumlah 19 untuk kelas lama dan 21 untuk kelas tidak lama yang akan dibagi menjadi data latih data uji. Dan menggunakan 30 data latih, kelas lama 19 dan kelas tidak lama 21.10 data uji kelas lama 5 dan kelas tidak lama 5. Hasil dari 10 data uji, terdapat 3 data yang hasil klasifikasi dengan mengaunkan metode Naïve Bayes yang tidak sama dengan hasil data yang sebenarnya. Keakurasian ketepatan hasil klasifikasi dengan naïve bayes untuk mengetahui tingkat kecenderungan penyelesaian studi mahasiswa sistem memiliki akurasi kebenaran sistem sebesar 70% dan nilai error sebesar 30%.

2. Penelitian dilakukan oleh Meinggian Vilian Sari, lulusan Universitas Muhammadiyah Gresik Tahun 2014. Penelitian untuk “*Memprediksi Prestasi (IPK) Mahasiswa Berdasarkan Latarbelakang Sekolah Asal Dana Tribut Mahasiswa ketika Awal Masuk Kuliah Menggunakan Naïve Bayes*”. Adapun data yang diambil dalam penelitian ini adalah sampel dari 103 *record* dengan kelas “Tinggi” dan “Rendah” masing-masing berjumlah 69 dan 34 yang akan dibagi menjadi data latih data uji. Dalam penelitiannya menggunakan 6 variable, adapun variabel yang dipakai : Instansi Sekolah, Status Sekolah, Jurusan Sekolah, Motivasi Pilihan Kuliah, Status Kerja dan Nilai Danem. Keakurasian hasil penelitian menggunakan metode Naïve Bayes ini menunjukkan akurasi tertinggi pada pengujian pertama adalah 84.62%. Dan pada pengujian keempat, ketiga percobaan menghasilkan nilai sensitivitas 100%, yang artinya semua data uji kelas tinggi diprediksi secara benar mempunyai kelas tinggi.
3. Penelitian dilakukan oleh Mohammad Zoqi Sarwani, Magister Ilmu Komputer, Universitas Brawijaya. Dalam acara Seminar Nasional Sistem Informasi Indonesia (SESINDO) di Institut Teknologi Sepuluh Nopember (ITS) Surabaya Tahun 2015. Penelitiannya berjudul : “*Analisis Twitter untuk Mengetahui Karakter Seseorang Menggunakan Algoritma Naïve Bayes Classifier*”. Dalam penelitian ini dilakukan analisis terhadap *tweets* atau *posting* kata yang terdapat pada Twitter yang dijadikan sebagai atribut. Kelas kepribadian yang digunakan berdasarkan konsep MBTI (Myer-Briggs Type Indicator), dari 16 indikator yang kemudian dikelompokkan menjadi 4 kelas tipe kepribadian yaitu : “Guardian”, “Artisan”, “Idealis” dan “Rasional”. Proses klasifikasi menggunakan algoritma Naïve Bayes Classifier. Pada uji coba yang dilakukan dalam penelitian ini, peneliti menggunakan 10 pengguna Twitter sebagai data latih dan 10 pengguna Twitter sebagai data uji. Hasil yang diperoleh tersebut kemudian dibandingkan dengan data dari pakar, sehingga didapat keakurasian data mencapai 100%.