



INSTITUT TEKNOLOGI DEL

**Prediksi *Turnover* Karyawan Menggunakan Algoritma KNN dan Algoritma
Naive Bayes**

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar Diploma Teknik

11315004	Sari Three Mayssi Siahaan
11315005	Indra Vincentius Manik
11315009	Sariaman Situmorang

**FAKULTAS TEKNIK INFORMATIKA DAN ELEKTRO
PROGRAM STUDI TEKNIK INFORMATIKA
LAGUBOTI
AGUSTUS 2018**

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya penulis sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah penulis nyatakan dengan benar.

Nama : Sari Three Mayssi Siahaan

NIM : 11315004

Tanda Tangan :

Tanggal : 31 Agustus 2018

Nama : Indra Vincentius Manik

NIM : 11315005

Tanda Tangan :

Tanggal : 31 Agustus 2018

Nama : Sariaman Situmorang

NIM : 11315009

Tanda Tangan :

Tanggal : 31 Agustus 2018

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh :

1. Nama : Sari Three Mayssi Siahaan
NIM : 11315004
Program Studi : Diploma 3 Teknik Informatika
 2. Nama : Indra Vincentius Manik
NIM : 11315005
Program Studi : Diploma 3 Teknik Informatika
 3. Nama : Sariamman Situmorang
NIM : 11315009
Program Studi : Diploma 3 Teknik Informatika
- Judul Tugas Akhir : Prediksi *Turnover* Karyawan Menggunakan
Algoritma KNN dan Algoritma Naive Bayes

Telah berhasil dipertahankan di hadapan Dewan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Diploma Teknik, pada program studi Diploma 3 Teknik Informatika, Fakultas Teknik Informatika dan Elektro, Institut Teknologi Del.

DEWAN PENGUJI

- Pembimbing : Lit Malem Ginting, S. Si, MT. ()
- Penguji : Riyanthi Angrainy Sianturi, ()
S.Sos., M.Ds.
- Penguji : Ike Fitriyaningsih, S.Si., M.Si ()
- Ditetapkan di : Laguboti
- Tanggal : 31 Agustus 2018

KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmatNya, penulis dapat menyelesaikan Tugas Akhir ini. Penulisan Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Ahli Madya Fakultas Teknik Informatika dan Elektro Institut Teknologi Del. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Tugas Akhir ini, sangatlah sulit bagi penulis untuk menyelesaikan Tugas Akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

- 1) Bapak Lit Malem Ginting, S. Si, MT. selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan penulis dalam penyusunan Tugas Akhir ini;
- 2) Ibu Riyanthi Angrainy Sianturi, S.Sos., M.Ds.dan Ibu Ike Fitriyaningsih, S.Si., M.Si selaku dosen penguji yang telah memberikan kritik dan saran yang membangun dalam penyusunan Tugas Akhir ini;
- 3) Orangtua dan keluarga penulis yang telah memberikan bantuan dukungan material dan moral; dan
- 4) Teman dan sahabat yang telah banyak membantu dan mendukung penulis dalam menyelesaikan Tugas Akhir ini.

Akhir kata, penulis berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Tugas Akhir ini membawa manfaat bagi pengembangan ilmu.

Laguboti, 31 Agustus 2018

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Institut Teknologi Del, penulis yang bertanda tangan di bawah ini:

1. Nama : Sari Three Mayssi Siahaan
NIM : 11315004
Program Studi : Diploma 3 Teknik Informatika

2. Nama : Indra Vincentius Manik
NIM : 11315005
Program Studi : Diploma 3 Teknik Informatika

3. Nama : Sariaman Situmorang
NIM : 11315009
Program Studi : Diploma 3 Teknik Informatika

Fakultas : Teknik Informatika dan Elektro

Jenis Karya : Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Institut Teknologi Del **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty-Fee Right*) atas karya ilmiah penulis yang berjudul:

Prediksi *Turnover* Karyawan Menggunakan Algoritma KNN dan Algoritma Naive Bayes

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Institut Teknologi Del berhak menyimpan, mengalih/media-format dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir penulis selama tetap mencantumkan nama penulis sebagai penulis/pencipta dan sebagai pemilik Hak cipta.

Demikian pernyataan ini penulis buat dengan sebenarnya.

Dibuat di : Laguboti

Pada tanggal : 31 Agustus 2018

Yang menyatakan

(Sari Three Mayssi Siahaan) (Indra Vincentius Manik) (Sariaman Situmorang)

ABSTRAK

Nama : 1. Sari Three Mayssi Siahaan
2. Indra Vincentius Manik
3. Sariaman Situmorang

Prodi : Diploma 3 Teknik Informatika

Judul : Prediksi *Turnover* Karyawan Menggunakan Algoritma KNN dan Algoritma Naive Bayes

Turnover adalah berhentinya individu dari keanggotaan suatu organisasi atau perusahaan. Beberapa faktor yang menyebabkan terjadinya *turnover* karyawan adalah jumlah gaji yang diterima rendah, penilaian perusahaan terhadap kinerja karyawan adalah buruk, ketidakpuasan karyawan terhadap perusahaan. Misalnya karyawan sudah bekerja di perusahaan untuk waktu yang lama namun karyawan belum pernah mendapatkan promosi atau kenaikan pangkat. Penelitian ini bertujuan untuk membangun sebuah simulator yang dapat memprediksi karyawan yang akan keluar atau bertahan di sebuah perusahaan, berdasarkan penilaian dari 9 variabel yang ada. Variabel tersebut adalah *satisfaction level*, *last evaluation*, *number of project*, *average monthly hours*, *time spend company*, *work accident*, *promotion last 5 years*, and *salary*. Simulator ini dibangun dengan menggunakan algoritma KNN dan naive bayes. Algoritma KNN dan naive bayes diuji sebanyak lima kali menggunakan data test untuk mengetahui tingkat akurasi terbaik. Algoritma KNN menghasilkan akurasi 95.8 persen, sedangkan algoritma naive bayes menghasilkan tingkat akurasi sebesar 66.04 persen. Algoritma KNN dengan akurasi 95.8 persen ini digunakan untuk menjadi model prediksi yang digunakan dalam simulator. Model prediksi ini kemudian diuji menggunakan data tunggal untuk mengetahui prediksi benar atau salah. Hasil yang diperoleh dari penelitian ini adalah sebuah simulator yang mampu memprediksi data karyawan yang keluar atau bertahan pada perusahaan, dengan menggunakan data tunggal sebanyak 210 data untuk menguji simulator yang telah dibangun. Berdasarkan hasil pengujian data tunggal maka diperoleh prediksi data benar sebanyak 198 dengan akurasi 94.28 persen.

Kata Kunci : *Turnover*, prediksi, *k-nearest neighbors*, *naive bayes*.

ABSTRACT

Name : 1. Sari Three Mayssi Siahaan
2. Indra Vincentius Manik
3. Sariamman Situmorang

Major : Diploma 3 Informatic Engineering

Title : Prediksi Turnover Karyawan Menggunakan Algoritma KNN dan Algoritma Naive Bayes

Turnover was stopping individuals from membership of an organization or company. Some of the factors that led to employee turnover is the amount of the salary is low, the company's assessment against the performance of an employee is bad, the dissatisfaction of employees towards the company. For example employees already working in the company for a long time but employees haven't been getting promotions or promotion. This research aims to build a simulator that can predict the employees who will be out or survive well at the company, based on judgments of 9 variables that exist. The variable is the satisfaction level, last evaluation, number of projects, the average monthly hours, time spend company, work accident, promotion last 5 years, and salary. This simulator was built using the algorithm KNN and naive bayes. Naive bayes algorithm KNN and tested as much as five times the test data are used to find out the best accuracy. KNN algorithm produces the accuracy of 95.8 per cent, while the naive bayes algorithm generates a degree of accuracy of 66.04 percent. KNN algorithm with accuracy of 95.8 percent it used to be a model predictions used in the simulator. Model predictions are then tested using single data to know the predictions are right or wrong. The results obtained from this research is a simulator that is able to predict employee data that came out or hang on to the company, using a single data as much as 210 simulator to test data that has been built. Based on the test results of single data then the data correct prediction earned as many as 198 with accuracy 94.28 percent.

Keywords: Turnover, prediction, k-nearest neighbours, naive bayes

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
1.3 Lingkup	2
1.4 Pendekatan	3
1.5 Sistematika Penyajian	3
BAB 2 STUDI LITERATUR	5
2.6 Sumber Daya Manusia (<i>Human Resources</i>)	5
2.7 Turnover	5
2.8 Sumber Data	6
2.9 Evaluasi Variabel	6
2.4.1 <i>Independent</i> Variabel	7
2.4.2 <i>Dependent</i> Variabel	7
2.10 Machine Learning	7
2.11 Algoritma K-Nearest Neighbor (KNN)	9
2.12 Algoritma K-Fold Cross Validation	11
2.13 Algoritma Naïve Bayes	12
2.14 Bahasa pemrograman Python	12
2.15 Data Preprocessing	13
2.16 Kesimpulan	15
BAB 3 ANALISIS DAN DESAIN	17
3.1 Analisis Data	17
3.2 Analisis Parameter	19
3.2.1 Independent Variabel	19
3.2.2 Dependent Variabel	20
3.3 Data Collection	20
3.4 Analisis Algoritma KNN	22
3.5 Analisis Algoritma Naive Bayes	28
3.6 Analisis Data <i>Transformation</i>	30
3.7 Desain Eksperimen dengan Algoritma KNN	30
3.8 Desain Eksperimen dengan Algoritma Naïve Bayes	32
3.9 Desain Simulator	33
3.10 <i>Use Case Scenario</i>	35
BAB 4 ANALISIS DAN DESAIN	36
4.1 Implementasi	36
4.1.1 Skenario Implementasi	36
4.1.2 Lingkungan Implementasi	36
4.1.3 Batasan Implementasi	37
4.1.4 Pelaksanaan Implementasi	37
4.1.4 Proses Prediksi Pada Algoritma KNN	38
4.1.5 Proses Prediksi Pada Algoritma Naive Bayes	41

4.1.6	Proses Prediksi Data Tunggal.....	43
BAB 5 HASIL IMPLEMETASI DAN PEMBAHASAN		46
5.1	Hasil Implementasi dan Pembahasan	46
5.2	Hasil Implementasi Data Kolektif dengan Algoritma K-Nearst Neighbor.....	46
5.4	Hasil Implementasi Data Kolektif dengan Algoritma Naive Bayes.....	49
5.4	Hasil Implementasi Data Tunggal	51
BAB VI SARAN DAN KESIMPULAN		53
6.1	Kesimpulan	53
6.2	Saran	53
References		54
Lampiran.....		lvi

DAFTAR TABEL

Tabel 1. Independent Variabel	19
Tabel 2. Dependent Variabel	20
Tabel 3. Komposisi Data Collection	20
Tabel 4. Komposisi Data Kolektif	21
Tabel 5. Data <i>Training</i>	Error! Bookmark not defined.
Tabel 6. Data tes.....	Error! Bookmark not defined.
Tabel 7. Hasil Pengurangan X1-X2	23
Tabel 8. Hasil Perpangkatan	24
Tabel 9. Hasil Penjumlahan dan Pengakaran	25
Tabel 10. Pengurutan Ranking.....	25
Tabel 11. Urutan Lima Data	27
Tabel 12. Penyesuaian Class	28
Tabel 13. Data Percobaan 1 Naive Bayes	28
Tabel 14. Data Percobaan 2 Naive Bayes	29
Tabel 15. Use Case Scenario.....	35
Tabel 16. Hasil Prediksi dengan Algoritma KNN	47
Tabel 17. Komposisi Data Pengujian I KNN.....	47
Tabel 18. Komposisi Data Pengujian II KNN	47
Tabel 19. Komposisi Data Pengujian III KNN	48
Tabel 20. Komposisi Data Pengujian IV KNN.....	48
Tabel 21. Komposisi Data Pengujian V KNN	48
Tabel 22. Akurasi Tiap Pengujian dengan Algoritma Naive bayes	49
Tabel 23. Komposisi Data Pengujian I Naive Bayes	50
Tabel 24. Komposisi Data Pengujian II Naive Bayes.....	50
Tabel 25. Komposisi Data Pengujian III Naive Bayes	50
Tabel 26. Komposisi Data Pengujian IV Naive Bayes	51
Tabel 27. Komposisi Data Pengujian V Naive Bayes.....	51
Tabel 28. Detail Data Tunggal.....	52
Tabel 29. Komposisi Data Tunggal	52

DAFTAR GAMBAR

Gambar 1. Supervised Learning Algorithms	8
Gambar 2. Unsupervised Learning Algorithms	9
Gambar 3. Data Transformation	30
Gambar 4. Desain Eksperimen Algoritma KNN.....	31
Gambar 5. Desain Eksperimen dengan Algoritma Naïve Bayes	32
Gambar 6. Interface Simulator.....	33
Gambar 7. Data Flow Simulator	34
Gambar 8. Potongan Kode Untuk Import Data Set.....	38
Gambar 9. Potongan Kode untuk Pemisahan Data Training dan Data Set.....	38
Gambar 10. Potongan Kode untuk Klasifikasi.....	39
Gambar 11. Potongan Kode untuk Pengecekan.....	39
Gambar 12. Output Pengecekan.....	40
Gambar 13. Potongan Kode Penghitungan Akurasi	40
Gambar 14. Output Penghitungan Akurasi	40
Gambar 15. Import Data Set dengan Naive Bayes	41
Gambar 16. Pemisahan data Training dengan Data Test dengan Naive Bayes	41
Gambar 17. Proses Klasifikasi dengan Naive Bayes	41
Gambar 18. Potongan Kode untuk Pengecekan dengan Naive Bayes	42
Gambar 19. Output Pengecekan dengan Naive Bayes.....	42
Gambar 20. Penghitungan akurasi	42
Gambar 21. Output Penghitungan Akurasi	43
Gambar 22. Import Data Tunggal	43
Gambar 23. Pengenalan Nilai Salary	44
Gambar 24. Potongan Kode Mengubah Nilai Salary.....	44
Gambar 25. Potongan Kode Input Nilai.....	44
Gambar 26. Contoh Input Nilai.....	45
Gambar 27. Potongan Kode Pengenalan Model Data.....	45
Gambar 28. Potongan Kode Prediksi Tunggal.....	45
Gambar 29. Hasil Prediksi	45
Gambar 30. Hasil Implementasi k-NN.....	Error! Bookmark not defined.

Gambar 31. Hasil Implementasi Data Kolektif.....	49
--	----

BAB 1

PENDAHULUAN

Bab I ini berisi uraian latar belakang pengerjaan tugas akhir, tujuan yang ingin dicapai, lingkup pengerjaan tugas akhir, pendekatan yang dilakukan dan sistematika penyajian dokumen.

1.1 Latar Belakang

Sumber daya manusia adalah aset perusahaan yang paling utama dalam penelitian ini sumber daya manusia yang dimaksud adalah karyawan, dalam sebuah perusahaan kemampuan karyawan tidak dapat digandakan [4]. Sumber Daya Manusia (SDM) merupakan *resource* yang sangat penting bagi keberhasilan dan eksistensi suatu perusahaan untuk mencapai tujuan yang ditetapkan dalam persaingan dunia bisnis. Oleh karena itu, sudah menjadi suatu keharusan bahwa SDM yang dimiliki perusahaan dapat dikelola dan dijaga semangat kerjanya agar senantiasa memberikan dampak positif bagi perkembangan perusahaan tersebut [2]. Akan tetapi fakta yang terjadi adalah kinerja perusahaan yang sudah baik dapat rusak, baik itu secara langsung maupun tidak langsung oleh perilaku karyawan. Salah satu bentuk perilaku karyawan tersebut adalah intensi keluar (*turnover intention*) yang dapat berujung pada keputusan karyawan yang meninggalkan pekerjaannya (*turnover*). *Turnover* didefinisikan sebagai berhentinya individu dari keanggotaan suatu organisasi atau perusahaan [1]. Sedangkan *turnover intention* didefinisikan sebagai sejauh mana karyawan berencana meninggalkan organisasi atau perusahaan [2].

Turnover sangat berpengaruh terhadap kinerja dari perusahaan karena akan mengganggu pekerjaan yang telah dikerjakan oleh karyawan yang meninggalkan perusahaan. Karyawan yang meninggalkan perusahaan menyebabkan terganggunya keberlanjutan sebuah pekerjaan karena karyawan baru yang akan menangani pekerjaan tersebut membutuhkan waktu untuk mempelajari proses bisnis pekerjaan barunya. Terdapat beberapa faktor yang menyebabkan karyawan keluar dari sebuah perusahaan diantaranya adalah ketidakpuasan terhadap kerja yang diterima karyawan atau ketidakpuasan karyawan terhadap perusahaan atas jam kerja yang ditentukan, jumlah pekerjaan yang diterima, *salary* yang didapat oleh karyawan atau penilaian perusahaan terhadap pekerjaan karyawan [8].

Namun, tidak semua karyawan memilih untuk meninggalkan perusahaan, meskipun karyawan mengalami salah satu dari faktor-faktor terjadinya *turnover*. Ada karyawan yang memiliki gaji rendah namun tetap bertahan pada sebuah perusahaan hal ini dipengaruhi oleh faktor lain sehingga karyawan memilih untuk tetap bertahan pada sebuah perusahaan. Oleh karena itu perlu dilakukan penelitian untuk memprediksi apakah karyawan akan keluar dari satu perusahaan berdasarkan variabel tertentu. Untuk itu dibutuhkan model prediksi yang mampu memprediksi karyawan yang akan keluar dari suatu perusahaan berdasarkan faktor-faktor *turnover* sebagai variabel penentu untuk memprediksi terjadinya *turnover*. Algoritma yang diterapkan untuk membangun model prediksi tersebut adalah algoritma *K-Nearest Neighbor* (k-NN) dan *Naive Bayes*.

1.2 Tujuan

Tujuan yang ingin dicapai dari hasil pelaksanaan tugas akhir ini adalah :

1. Membandingkan antara algoritma KNN dan Naïve Bayes, kemudian memilih algoritma terbaik dengan capaian akurasi terbaik.
2. Membangun simulator prediksi *turnover* karyawan, menggunakan algoritma KNN berdasarkan variabel yang ada.

1.3 Lingkup

Dalam pelaksanaan Tugas Akhir ini, akan dilakukan pembahasan mengenai:

1. Data masukan yang akan digunakan adalah data yang diperoleh dari data *Human Resource Analytics* pada situs <https://www.kaggle.com>.
2. Algoritma pembelajaran yang digunakan yaitu algoritma *k-Nearest Neighbor* (KNN) dan *Naive Bayes*.
3. Untuk pembuatan simulator digunakan sebuah *tools* yaitu *Python 3.6*.
4. Variabel yang digunakan ada sepuluh, yaitu: *satisfaction*, *last evaluation*, *number of project*, *average monthly hours*, *time spend company*, *work accident*, *left*, *promotion last five years*, *job description*, dan *salary*.
5. Membandingkan algoritma *k-Nearest Neighbor* (KNN) dan *Naive Bayes* berdasarkan akurasi yang diperoleh dalam proses prediksi.

1.4 Pendekatan

Metodologi penelitian yang akan digunakan selama pelaksanaan Tugas Akhir meliputi:

1. Studi Literatur

Studi Literatur dilakukan dengan cara mengumpulkan bahan-bahan yang berhubungan dengan topik. Hasil yang diperoleh dengan pendekatan ini berupa studi literatur terkait *employee turnover* dan algoritma yang akan digunakan.

2. Analisis Data

Tahap analisis dilakukan dengan melakukan analisis terhadap parameter yang bersumber dari Kaggle.com, yaitu dengan mempelajari setiap parameter beserta nilai yang tertera pada data tersebut.

3. Mendesain Model Prediksi

Dalam tahap ini dilakukan perancangan model aplikasi prediksi untuk memprediksi karyawan yang keluar dari perusahaan tersebut.

4. Membangun Simulator

Setelah merancang model prediksi, tahap selanjutnya adalah pembangunan simulator.

5. Menguji Model Prediksi

Setelah membangun simulator, penulis akan melanjutkan ke tahap pengujian model prediksi yang telah diimplementasikan dalam sistem untuk mendapatkan akurasi model.

6. Analisis Hasil Prediksi

Hasil yang telah diperoleh pada proses pengujian model prediksi akan dianalisis untuk mendapatkan kesimpulan dari hasil pengujian model prediksi.

1.5 Sistematika Penyajian

Secara garis besar laporan Tugas Akhir ini dibagi dalam beberapa bab sebagai berikut :

1. Bab I Pendahuluan

Bab ini berisi uraian tentang latar belakang, tujuan, lingkup, pendekatan serta sistem aplikasi penyajian dalam mengerjakan Tugas Akhir.

2. Bab II Tinjauan Pustaka

Bab ini berisi rangkuman informasi, metodologi yang berkaitan dengan topik Tugas Akhir.

3. Bab III Analisis dan Desain

Bab ini berisi penjelasan singkat mengenai algoritma yang digunakan untuk pengerjaan Tugas Akhir.

4. Bab IV Implementasi dan Pembahasan

Bab ini berisi penjelasan bagaimana kegiatan-kegiatan yang dilakukan dalam implementasi aplikasi dan hasil yang diperoleh dari pengembangan aplikasi pada pelaksanaan Tugas Akhir

5. Bab V Kesimpulan dan Saran

Bab ini berisi kesimpulan dan saran sesudah Tugas Akhir yang dilakukan.

BAB 2

STUDI LITERATUR

Pada bab ini, dijelaskan mengenai *study literature* dan algoritma yang akan digunakan untuk pembuatan model prediksi *turnover* karyawan.

2.6 Sumber Daya Manusia (*Human Resources*)

Sumber Daya Manusia (SDM) adalah salah satu faktor yang mendukung keberhasilan sebuah organisasi. Oleh karena itu Sumber Daya Manusia harus dikelola dengan baik agar tujuan sebuah organisasi dapat tercapai. Dalam Dunia bisnis Sumber Daya Manusia yang memiliki kemampuan dan dapat diandalkan sangat dibutuhkan. Dalam sebuah organisasi dibutuhkan kerjasama dari Sumber Daya Manusia yang ada untuk mencapai tujuan. Tujuan yang dimaksud adalah visi dan misi yang sudah diatur dan disetujui bersama. Agar hubungan kerjasama dapat terlaksana dengan baik maka perusahaan harus meningkatkan kesejahteraan Sumber Daya Manusia yang ada pada perusahaan. Dalam penelitian ini Sumber Daya Manusia yang dimaksud adalah karyawan. Karyawan merupakan orang yang bekerja kepada orang lain, orang yang menjual jasa, waktu dan tenaganya dan mendapatkan kompensasi dari kinerja yang telah dicapai. Kinerja karyawan merupakan hasil kerja yang telah dicapai sesuai dengan tanggung jawab yang telah diberikan baik secara kualitas maupun kuantitas. Peranan karyawan pada sebuah perusahaan adalah ikut serta dalam perencanaan, sistem, proses dan pencapaian dari tujuan yang akan dicapai perusahaan[20].

2.7 Turnover

Turnover adalah masalah yang terus menerus terjadi dalam organisasi dan hal ini biasa terjadi pada setiap jenis organisasi dan pada setiap tingkat atau divisi pada organisasi. *Turnover* didefinisikan sebagai berhentinya individu dari keanggotaan suatu organisasi atau perusahaan [1]. *Employee turnover* mengarah pada kenyataan akhir yang dihadapi perusahaan (kehilangan sejumlah karyawan) pada periode tertentu, sedangkan keinginan pindah kerja (*turnover intentions*) sendiri mengacu kepada hasil evaluasi individu mengenai

kelanjutan hubungannya dengan sebuah perusahaan yang belum diwujudkan dalam tindakan nyata meninggalkan perusahaan tersebut [12].

Employee turnover dapat berakibat buruk pada perusahaan karena pada saat pergantian karyawan terjadi, karyawan yang tinggal akan kesulitan untuk melanjutkan pekerjaan yang ditinggalkan oleh karyawan yang pergi, selain itu pihak perusahaan juga akan membutuhkan biaya untuk melakukan perekrutan karyawan baru, dan akan menghabiskan waktu untuk melakukan perekrutan karyawan baru dan melakukan pelatihan terhadap karyawan yang baru [13].

Untuk memperkecil terjadinya *employee turnover* strategi yang dapat digunakan adalah membuat surat perjanjian atau kontrak kerja antara perusahaan dengan karyawan, yang berisi komitmen karyawan terhadap pekerjaan dan lama waktu untuk bekerja pada sebuah perusahaan. Dengan adanya kontrak kerja diharapkan karyawan lebih termotivasi untuk bekerja dan bertahan pada sebuah perusahaan sampai pada waktu yang ditentukan pada kontrak kerja yang telah disepakati dan disetujui bersama, strategi lain yang dapat digunakan adalah perusahaan menyediakan akses kepada karyawan untuk menambah pengetahuan dan untuk mendapatkan informasi terkait pekerjaan yang ditanggungjawab oleh karyawan [13].

2.8 Sumber Data

Sumber data yang digunakan untuk membuat model prediksi adalah Kaggle.com. Kaggle.com adalah salah satu komunitas untuk data *scientists*, *statisticians*, dan *machine learning*. Orang-orang yang masuk dalam kaggle pada umumnya menunjukkan kemampuan mereka untuk menyelesaikan permasalahan-permasalahan yang ada pada dunia industri.

2.9 Evaluasi Variabel

Variabel dalam penelitian ini dapat didefinisikan sebagai parameter yang akan digunakan dalam mempengaruhi hasil akhir dari penelitian. Variabel ini nantinya akan menjadi salah satu tolak ukur dalam penelitian. Variabel dalam penelitian ini dibagi terhadap 2 kelompok berdasarkan kebutuhannya, yaitu *Independent variabel* dan *dependent variabel*.

2.4.1 *Independent Variabel*

Independent variabel atau *variabel bebas* adalah variabel yang mempengaruhi atau sebab perubahan timbulnya variabel terikat (*dependen*). Variabel independen disebut juga dengan variabel perlakuan, klausa, risiko, variabel stimulus, *antecedent*, variabel pengaruh, *treatment*, dan variabel bebas. Dapat dikatakan variabel bebas karena dapat mempengaruhi variabel lainnya. *Independent variabel* merupakan *variabel* yang akan mempengaruhi hasil dari prediksi.

2.4.2 *Dependent Variabel*

Dependent variabel adalah variabel yang dipengaruhi, akibat dari adanya variabel bebas. *variabel dependent* disebut juga dengan variabel terikat, *variabel output*, Konsekuen, variabel tergantung, kriteria, variabel terpengaruh, dan variabel efek. Dikatakan sebagai variabel terikat karena variabel terikat dipengaruhi oleh variabel independen (*variabel bebas*). *Dependent variabel* merupakan variabel yang akan diprediksi.

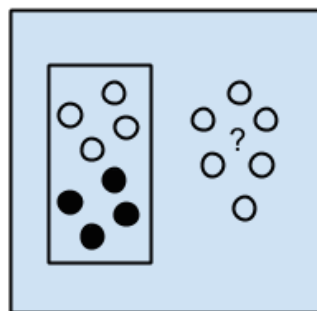
2.10 Machine Learning

Machine learning adalah kecerdasan buatan untuk memprediksi kejadian yang akan datang. Menurut Arthur Samuel *machine learning* diartikan sebagai bidang pelajaran yang memberikan komputer kemampuan untuk belajar tanpa harus diprogram secara eksplisit [6]. *Machine learning* merupakan bagian dari kecerdasan buatan atau *Artificial Intelligence* (AI) yang menggunakan perhitungan komputasi, dengan cara mendesain sistem, agar dapat mempelajari data. Sistem dapat belajar dan meningkatkan kemampuan dari waktu ke waktu untuk menyempurnakan model yang dapat digunakan untuk memprediksi hasil berdasarkan data yang dipelajari sebelumnya [11]. Pembelajaran mesin merupakan proses pembangunan model setelah menemukan pengetahuan dari sebuah data atau sekumpulan data. Secara umum pembelajaran mesin ini dianggap sebagai proses penerapan sumber daya berbasis *computer* untuk menerapkan algoritma belajar [6]. Pendekatan dalam *Machine Learning* dibagi menjadi dua yaitu *inductive learning* dan *deductive learning*. *Deductive learning* adalah proses belajar berdasarkan informasi

yang *general* untuk menyelesaikan masalah yang spesifik. Sedangkan *Inductive learning* adalah proses belajar berdasarkan informasi yang spesifik seperti contoh kasus yang sudah diselesaikan beserta hasilnya guna mendapatkan pola dalam menyelesaikan masalah, dan pola yang didapatkan dari informasi spesifik digunakan untuk menyelesaikan masalah yang baru. Pembelajaran mesin dibagi kedalam dua jenis yaitu *supervised learning* dan *unsupervised learning*.

1. *Supervised Learning*

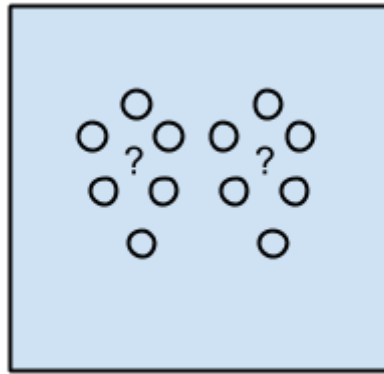
Supervised Learning (pembelajaran terarah) adalah sebuah pendekatan dimana sudah terdapat data yang dilatih (*data training*), dan terdapat *variabel* yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan suatu data ke data yang sudah ada. Pada *supervised learning* suatu model disiapkan melalui suatu proses pelatihan untuk membuat prediksi dan dikoreksi ketika prediksi tersebut salah. Proses pelatihan dilakukan secara berulang sampai model mencapai tingkat akurasi yang diinginkan pada data pelatihan. Pemodelan *supervised learning* dapat dilihat pada Gambar 1. Supervised Learning Algorithms Contoh dari *supervised learning* adalah klasifikasi dan regresi. Klasifikasi terdiri atas *K-Nearest Neighbor*, *Support Vector Machine*, *Naive Bayes*, dan *Bayesian Network*, sedangkan regresi terdiri dari *Linear-Regression*, *Non-Linear Regression*, dan *Support Vector Regression*.



Gambar 1. Supervised Learning Algorithms

2. *Unsupervised Learning*

Unsupervised learning (pembelajaran tak terarah) tidak memiliki data yang dilatih (data *training*), sehingga dari data yang ada, kita mengelompokkan data tersebut menjadi 2 bagian atau 3 bagian dan seterusnya [14]. Contoh pemodelan *unsupervised learning* dapat dilihat pada Gambar 2. Unsupervised Learning Algorithms Penggolongan *Unsupervised Learning* adalah *clustering* dan *Assosiation Leraning*. *Clustering* terdiri dari *K-Means*, *Hierarchical Clustering*, dan *Gaussian Mixture*, sedangkan *Assosiation Learning* terdiri dari *FP-Growth*, *Apriori*, dan *DBScan*



Gambar 2. Unsupervised Learning Algorithms

Bahasa pemrograman yang digunakan untuk *menchine learning* adalah python, java, C++,C, javascript, scala, dan julia. Sementara tools yang digunakan terbagi dua yaitu untuk coding dan simplified, yang tergolong pada coding adalah K-Keras, Spahark H2O, Scikit learn, Tensor flow, dan toools yang tergolong pada simplified adalah orange, Weka dan rapidminer .

2.11 Algoritma K-Nearest Neighbor (KNN)

Algoritma *k-Nearst Neigbor(KNN)* adalah metode pada machine learning metode ini termasuk dalam metode *supervised*. Algoritma ini dilakukan dengan cara menentukan mayoritas dari K. K adalah objek pada data *training* yang paling dekat pada data baru atau data testing yang biasa dikatakan sebagai tetangga terdekat.[16]

Tujuan algoritma k-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan *training sample*. Klasifikasi dilakukan dengan cara memilih data terbanyak diantara klasifikasi dari k objek. Algoritma k-NN bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk

menentukan k-NN-nya. Algoritma k-NN menggunakan *neighborhood classification* sebagai nilai prediksi dari nilai instace yang baru.

Kelebihan algoritma k-NN adalah mampu memodelkan fungsi tujuan yang kompleks dengan sejumlah perkiraan kompleks lokal, selain itu informasi yang tersimpan pada data latih tidak pernah hilang. k-NN merupakan teknik klasifikasi yang sederhana, tetapi mempunyai hasil kerja yang cukup bagus. k-NN telah memasuki *top ten* algoritma *data mining*.

Kelemahan utama pada k-NN adalah harus menyimpan dalam jumlah dan karakteristik yang sangat besar pada data latih. Karena k-NN mendasarkan pada perhitungan jarak antara data uji dengan data latih maka k-NN membutuhkan waktu yang sangat besar pada saat proses prediksi. Disamping itu, data latih juga memberikan pengaruh yang besar pada target keluaran. *Noise* pada data juga dapat mengurangi kinerja k-NN, sehingga data dan noise yang tidak penting harus dieleminasi sehingga waktu komputasi dan *error* dapat dikurangi. [15].

Secara manual langkah-langkah untuk menghitung metode k-Nearest Neighbor adalah sebagai berikut:

1. Menentukan parameter K

Dalam algoritma k-NN data baru harus diklasifikasikan berdasarkan jarak data baru dengan tingkat kemiripan data baru terdekat terhadap pola, jumlah data terdekat ditentukan dengan nilai K. Nilai k dapat diperoleh dengan menggunakan algoritma *K-Fold Cross Validation* yaitu dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut data *input* yang di acak[7].

Untuk melakukan pengujian terhadap akurasi algoritma yang digunakan dapat menggunakan persamaan 1 berikut

$$Akurasi = \frac{jumlah\ klasifikasi\ benar}{Jumlah\ data\ uji} \times 100\%$$

(1)

2. Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan.

Adapun Rumus persamaan dari algoritma K-Nearest Neighbor untuk menghitung jarak adalah pada persamaan 2 berikut :

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

(2)

Dimana:

- a. x_1 = Sampel data
 - b. x_2 = Data uji
 - c. i = Variabel data
 - d. d = Jarak
 - e. p = Dimensi data
3. Mengurutkan jarak yang telah dibentuk
 4. Menentukan jarak terdekat sampai K
 5. Memasangkan kelas yang bersesuaian
 6. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut

2.12 Algoritma K-Fold Cross Validation

K-Fold Cross validation adalah algoritma yang dirancang untuk memberikan perkiraan akurat tanpa membuang membuang data yang terlalu banyak. *K-fold cross validation* sering digunakan untuk pemilihan model (parameter *tuning*), setelah parameter terbaik dipilih, algoritma ini akan dilatih dengan parameter terbaik yang telah dipilih pada seluruh data *training* dan data *test*.

Pengujian K-Fold Cross Validation dilakukan dengan cara bervariasi data yang digunakan untuk data *train* dan data *test*. Untuk mencari kekonsistenan kinerja sistem klasifikasi dilakukan pengujian sebanyak 5 kali, data *train* dan data *test* yang akan digunakan adalah data pada pengujian terakhir [23].

2.13 Algoritma Naïve Bayes

Naive Bayes adalah algoritma *supervised learning* yang sangat sederhana. Algoritma *naive bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori paling tepat. Secara umum formula *naive bayes* ada pada persamaan 3.

$$P(H|X) = \frac{P(X|H)XP(H)}{P(X)} \quad (3)$$

Dimana:

1. $P(H|X)$ adalah probabilitas akhir bersyarat (*posterior probability*) suatu hipotesis H terjadi jika diberikan bukti E terjadi.
2. $P(X|H)$ adalah probabilitas sebuah bukti E terjadi akan memengaruhi hipotesis H.
3. $P(H)$ adalah probabilitas awal (*prior probability*) hipotesis H terjadi tanpa memandang bukti apapun.
4. $P(X)$ adalah probabilitas awal (*prior probability*) bukti E tanpa memandang hipotesis atau bukti yang lain.

Tahapan dari proses algoritma *naive bayes* adalah :

1. Menghitung jumlah kelas/label.
2. Menghitung jumlah kasus per kelas.
3. Mengalikan semua variabel kelas.
4. Membandingkan hasil per kelas.

Pada proses klasifikasi algoritma *naive bayes* memiliki kekurangan dan kelebihan. Untuk kelebihan dari algoritma *naive bayes* adalah proses klasifikasi mudah untuk dibuat dan hasil yang diperoleh bagus. Sedangkan kekurangan dari algoritma *naive bayes* adalah Asumsi independence antar atribut membuat akurasi berkurang disebabkan oleh adanya keterikatan.

2.14 Bahasa pemograman Python

Python merupakan bahasa pemograman yang dirancang untuk kemudahan dalam pemograman computer dan fungsi. Salah satu tujuan desain dari bahasa

python adalah kode yang ada mudah untuk dimengerti karena sintaks kode nya. Bahasa *python* memiliki sintaks tertentu(bentuk) dan semantic yaitu memungkinkan untuk mengoreksi manipulasi data dan perhitungan yang dilakukan oleh computer dan memiliki library yang lebih [5]. *Python* diimplementasikan pertama kali pada tahun 1989 oleh Guido van Rossum di CWI (Institusi Penelitian Nasional di belanda) sebagai seorang suksesor untuk bahasa pemograman ABC (sebuah bahasa yang populer dan menjadi motivasi untuk mengembangkan bahasa *python*). Van Rossum adalah sebagai pengembang bahasa *python* yang utama dan perannya yang tengah terus dalam menentukan arah bahasa *Python*. *Python* digunakan dalam machine learning karena bahasa pemograman tidak langsung menggunakan *host* cpu saat melakukan *compile* tapi ditangani oleh *compiler* sehingga tidak terlalu memberatkan saat program melakukan *compile* [5].

Kelebihan Bahasa pemograman *python*

1. Pengembangan oleh pihak ketiga

The python Package Index(PyPI) konten angka dari Bahasa *python* yang dikembangkan untuk keperluan berbagai platforms dan dapat dihubungkan dengan Bahasa pemograman yang lain

2. Object-oriented

Python dapat juga digunakan untuk pemograman yang menggunakan prinsip berorientasi objek.

3. *User-friendly* Data Structure

Python memiliki daftar *built-in* dan *library* stuktur data yang dapat digunakan untuk membuat sebuah struktur data dengan runtime yang cepat.

Kekurangan bahasa pemograman *python* adalah *Speed can be an issue* maksudnya adalah kecepatan menjadi masalah, saat melakukan *compile code program*, proses nya sering menjadi lambat dari bahasa yang lain[17].

2.15 Data Preprocessing

Data *preprocessing* adalah tahap awal proses data *mining*. Sebelum melakukan pengujian terhadap algoritma yang pertama kali dilakukan adalah mempersiapkan data yang akan diperlukan. Tahap *preprocessing* mencakup

semua aktivitas, dan proses untuk mempersiapkan data yang akan digunakan. *Preprocessing* berfungsi untuk mentransformasikan data mentah atau mengubah data kedalam format tertentu sehingga siap untuk dimasukkan kedalam algoritma pengujian. Data *preprocessing* perlu dilakukan karena data yang digunakan belum baik, hal ini dikarenakan beberapa hal, diantaranya adalah

1. Incomplete

Incomplete adalah adanya kekurangan pada nilai-nilai atribut atau atribut tertentu.

2. Noisy

Data yang digunakan mengandung kesalahan atau nilai-nilai *outlier* yang tidak sesuai dengan harapan.

3. Inconsisten

Inconsisten adalah adanya ketidakcocokan dengan kode atau nama pada data dengan kode yang digunakan. Untuk mendapatkan data yang baik dibutuhkan integrasi kualitas data yang konsisten.

Metode yang digunakan pada data *preprocessing* adalah:

1. Data Cleaning

Data cleaning adalah proses pembersihan data. Pembersihan data yang dimaksud adalah mendeteksi perbedaan berupa faktor opsional, termasuk bentuk entri data yang tidak dirancang dengan baik, dan kesalahan yang sengaja dibuat. Beberapa hal yang dilakukan pada *data cleaning* yaitu mengisi nilai yang hilang (*missing values*) pada *fields*, mengidentifikasi *ouliers*, *smooth out noisy data*, memperbaiki *inconsistent data*, dan memisahkan *redundancy* yang diakibatkan *data integration*. Pada proses *data cleaning* juga dilakukan *filtering*, yaitu *Stopword Removal* yaitu melakukan penghapusan tanda baca pada teks yang digunakan pada klasifikasi.

2. Data Integrasi

Pada *data integrastion* dilakukan penggabungan data dari berbagai sumber, seperti *database*, dan *file* lainnya yang dapat mengakibatkan *redundancy*. Dengan metode data integrasi *redudancy* pada data dapat dideteksi. Beberapa hal yang dilakukan untuk mengatasi *redudancy* pada *data integration* adalah mengidentifikasi permasalahan *entity*, analisis *redudancy* dan *correlation*, *tuple duplication*, dan mendeteksi data *value conflict*.

3. Data Transformasi

Pada metode ini data di transformasikan sehingga proses *mining* menjadi lebih efisien dan bentuk yang ditemukan lebih mudah untuk dipahami. Strategi yang digunakan pada *transformation data* adalah *smoothing*, *atribut construction*, *aggregation*, *normalization*, *discretization*, *concept hierarchy generation for nominal data* [19]

4. Data Reduksi

Data reduction dilakukan untuk menghasilkan *data set* yang telah dikurangi dengan volume yang lebih kecil namun masih memiliki hasil analisis yang sama sebelum melakukan reduksi pada data. Adapun strategy yang digunakan pada *data reduction* adalah *dimensionality reduction*, *numerosity reduction* dan *data compression* [19]. Dalam penelitian ini metode yang digunakan pada proses *preprocessing* adalah data transformasi, hal ini dilakukan karena adanya salah satu nilai dari variabel yang ada pada dataset yang bertipe *string*(huruf) yaitu *sallary* jenis dari nilai dari variabel ini adalah *low*, *medium*, dan *high*. Variabel ini merupakan tingkatan level gaji yang diperoleh karyawan dari perusahaan. Untuk dapat memproses nilai dari variabel ini dalam proses prediksi, nilai dari variabel ini terlebih dulu harus dirubah menjadi kode dalam bentuk angka sehingga variabel ini dapat digunakan dalam proses prediksi.

2.16 Kesimpulan

Setelah melakukan pembelajaran terhadap bahan pustaka dan studi yang akan dikerjakan dari hal tersebut dapat diperoleh :

1. *Turnover* didefinisikan sebagai berhentinya individu dari keanggotaan suatu organisasi atau perusahaan.
2. Variabel yang digunakan dalam penelitian ini dibagi menjadi dua, yaitu variabel independent dan variabel dependent.
3. *Machine learning* berperan dalam mengenali, mengidentifikasi, ataupun memprediksi data tertentu dengan mempelajari data *History (Experience Data)*.
4. Algoritma k-NN dan Naive Bayes dapat digunakan untuk melakukan sebuah prediksi.
5. Sebelum melakukan pengujian prediksi dengan algoritma terlebih dahulu melakukan *preprocessing*.
6. Metode *preprocessing* yang digunakan adalah metode data transformasi.

BAB 3

ANALISIS DAN DESAIN

Pada bab ini menjelaskan analisis data yang digunakan, analisis sumber data dan analisis pembentukan pohon keputusan. Penjelasan dari setiap parameter adalah sebagai berikut:

3.1 Analisis Data

Dalam penelitian ini sumber data yang digunakan peneliti adalah dari Kaggle.com. Variabel yang digunakan penulis sebagai objek kajian yaitu *satisfaction level*, *last evaluation*, *number of project*, *average montly hours*, *time spend company*, *work accident*, *promotion last 5 years*, *job description*, *salary* dan *left*.

1. *Satisfaction Level*

Satisfaction level merupakan variabel yang mempresentasikan tingkat kepuasan perusahaan atas kinerja karyawan dengan angka. *Variabel* ini mempunyai nilai dari 0 sampai dengan 1, dengan nilai minimal adalah 0.09, nilai maximal 1 dengan *standart deviasi* 0.2486307, nilai rata-rata 0.6128335 dan median 0.64. Contoh: 0.09, 0.1, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, ..., 1).

2. *Last Evaluation*

Last Evaluation merupakan variabel yang menunjukkan penilaian perusahaan terhadap hasil evaluasi perusahaan atas kinerja karyawan dalam satu tahun. *Variabel* ini mempunyai nilai yang dimulai dari 0 sampai dengan 1 dengan nilai minimal 0.36, nilai maximal 1, *standart deviasi* 0.72, dengan rata-rata 0.71610174 dan nilai median adalah 0.72. Contoh: 0.36, 0.37, 0.38, 0.39, 0.40, ..., 1).

3. *Number of Project*

Number of Project merupakan variabel yang menunjukkan jumlah pekerjaan yang telah diselesaikan oleh karyawan selama bekerja di perusahaan.

Variabel ini mempunyai nilai bilangan bulat positif. Jumlah minimal proyek yang dikerjakan karyawan adalah 2 proyek, dan jumlah proyek maksimal yang dikerjakan adalah 7 median proyek yang dikerjakan karyawan adalah 4 proyek.

4. *Average monthly hour*

Average montly hour merupakan variabel yang menunjukkan rata-rata jam kerja karyawan perbulan minimum adalah 96 jam, dan jam kerja maksimal adalah 310 jam, dengan median 200 jam.

5. *Time spend company*

Time spend company merupakan variabel yang memberikan informasi kontrak kerja karyawan dengan perusahaan, dengan satuan tahun. Minimal kontrak kerja yang dijalin perusahaan dengan karyawan adalah 2 tahun, dan kontrak kerja maksimal yang dijalin perusahaan dengan karyawan adalah 10 tahun dengan median 3 tahun.

6. *Work Accident*

Work Accident adalah keadaan dimana apakah karyawan tersebut pernah mengalami kecelakaan kerja atau tidak. Nilai yang digunakan yaitu 0 atau 1, dimana 0 menyatakan karyawan tidak pernah mengalami kecelakaan kerja dan 1 menyatakan karyawan pernah mengalami kecelakaan kerja.

7. *Left*

Left adalah keadaan dimana karyawan meninggalkan perusahaan atau tidak. Nilai yang digunakan yaitu 0 atau 1, dimana 0 meyatakan karyawan masih tetap di perusahaan dan 1 menyatakan karyawan telah meninggalkan perusahaan.

8. *Promotion Last 5 Years*

Promotion Last 5 Years merupakan variabel yang menjelaskan apakah karyawan telah menerima Promosi oleh perusahaan dalam 5 tahun terakhir. Nilai yang digunakan dalam variabel ini adalah 0 atau 1, 0 menjelaskan bahwa karyawan belum menerima promosi dalam 5 tahun dan 1

menunjukkan karyawan telah menerima promosi dari perusahaan dalam 5 tahun.

9. *Job Description*

Job Description adalah menunjukkan karyawan tersebut bekerja di department mana dan *job description* ini memiliki nilai yaitu IT, management, sales, human resources, HR, product manager, marketing, accounting, techinal, support.

10. *Salary*

Salary adalah variabel yang memberikan informasi tingkatan gaji yang diperoleh karyawan, variabel ini memiliki yaitu: *low*, *medium*, dan *high*.

3.2 Analisis Parameter

Parameter yang digunakan dalam penelitian ini dibagi menjadi dua yaitu *independent* variabel dan *dependent* variabel.

3.2.1 Independent Variabel

Pada Tabel 1. Independent Variabel dapat dilihat parameter yang dijadikan sebagai *independent* variabel. *Independet variabel* atau *variabel* bebas adalah variabel yang mempengaruhi atau sebab perubahan timbulnya variabel terikat

Tabel 1. Independent Variabel

No.	Nama Variabel	Tipe Data
1.	<i>Satisfaction Level</i>	<i>Numeric</i>
2.	<i>Last Evaluation</i>	<i>Numeric</i>
3.	<i>Number of Project</i>	<i>Numeric</i>
4.	<i>Average montly hours</i>	<i>Numeric</i>
5.	<i>Time spend company</i>	<i>Numeric</i>
6.	<i>Work Accident</i>	<i>Numeric</i>
7.	<i>Promotion Last 5 Years</i>	<i>Numeric</i>

8.	<i>Job Description</i>	<i>String</i>
9.	<i>Salary</i>	<i>String</i>

3.2.2 Dependent Variabel

Dependent variabel adalah variabel yang dipengaruhi, akibat dari adanya variabel bebas. Dikatakan sebagai variabel terikat karena variabel terikat dipengaruhi oleh variabel independen (variabel bebas). *Dependent variabel* yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2.

Dependent Variabel

Tabel 2. Dependent Variabel

No.	Nama Variabel	Tipe Data
1.	<i>Left</i>	Int

3.3 Data Collection

Pada penelitian ini data *collection* yang digunakan adalah sebesar 14.999, sebelum data ini digunakan untuk proses prediksi terlebih dahulu data dibagi menjadi dua bagian yaitu data kolektif dan data tunggal. Komposisi data dapat dilihat pada Tabel 3. Komposisi Data Collection

Tabel 3. Komposisi Data Collection

No	Data Collection	Jumlah	Persentase
1	Data Kolektif	14.789	87 %
2	Data Tunggal	210	13 %
Total		14.999	100%

Data kolektif tersebut akan dibagi menjadi dua bagian yaitu data *training* dan data *test*. Sehingga data *collection* yang digunakan pada penelitian ini adalah sebagai berikut:

1. Data Kolektif

Data kolektif adalah data yang digunakan untuk prediksi *turnover* karyawan pada proses prediksi secara kolektif. Jumlah data kolektif yang digunakan dalam penelitian ini adalah 87 % dari data set yaitu 14.789 data karyawan, data kolektif ini terbagi menjadi dua bagian yaitu:

a. Data Training

Training data set merupakan data yang digunakan oleh algoritma klasifikasi untuk membentuk sebuah *model classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk memprediksi kelas data baru yang belum pernah ada. Dalam penelitian ini yang data *training* adalah 80 % dari seluruh data kolektif yang akan digunakan. Adapun jumlah data yang digunakan pada penelitian ini berjumlah 14.789. Sehingga yang menjadi data *training* adalah 11.831 data karyawan.

b. Data Test

Data *test* merupakan data yang digunakan untuk mengukur sejauh mana *Classifier model* berhasil melakukan klasifikasi dengan benar. Data *test* yang digunakan tidak boleh terdapat pada data *training*. Dalam penelitian yang menjadi data *test* adalah seluruh data kolektif yang tidak digunakan pada data *training*. Jumlah data yang tidak digunakan pada data *training* adalah 20 % yaitu 2958 data karyawan. Secara detail komposisi data kolektif dapat dilihat pada Tabel 4. Komposisi Data Kolektif

Tabel 4. Komposisi Data Kolektif

No	Data Kolektif	Jumlah	Persentase
1	Data Training	11.831	80 %
2	Data Test	2.958	20 %
Total		14.789	100 %

2. Data Tunggal

Data tunggal adalah data yang belum tersusun atau dikelompokkan kedalam kelas interval Data tunggal ini digunakan sebagai data untuk mengecek simulator yang dibangun, dengan memasukkan data secara satu persatu kedalam simulator. Tabel data tunggal data dilihat pada lampiran.

3.4 Analisis Algoritma KNN

Algoritma k-Nearest Neighbor adalah algoritma *supervised learning*. Algoritma ini bekerja dengan cara mengklasifikasikan berdasarkan jumlah mayoritas kategori K-tetangga terdekat. K-NN banyak digunakan dalam aplikasi data mining, *statistical pattern recognition*, *image processing* dan lain-lain. Algoritma ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel-sampel dari data *training*. Pada penelitian ini contoh kasus yaitu tugas akhir ini, peneliti menginginkan untuk melakukan prediksi *turnover* karyawan menggunakan algoritma k-NN. Berikut ini adalah hasil perhitungan manual dari data *turnover* karyawan menggunakan algoritma *k-nearest neighbour*. Data *training* dapat dilihat pada Tabel 5.

Tabel 5. Data Taraining

No	Satisfaction Level	Last Evaluation	Number Project	Average Montly hours	Time Spend company	Work accident	Left	Promotion Last 5 years	Job Description	Salary
1	0,38	0,53	2	157	3	0	1	0	Sales	Low
2	0,8	0,86	5	262	6	0	1	0	Sales	Medium
3	0,11	0,88	7	272	4	0	1	0	Sales	Medium
4	0,72	0,87	5	223	5	0	1	0	Sales	Low
5	0,37	0,52	2	159	3	0	1	0	Sales	Low
6	0,58	0,74	4	215	3	0	0	0	Sales	Low
7	0,82	0,67	2	202	3	0	0	0	Sales	Low
8	0,45	0,69	5	193	3	0	0	0	Sales	Low
9	0,78	0,82	5	247	3	0	0	0	Sales	Low
10	0,49	0,6	3	214	2	0	0	0	Sales	Low

Tabel 6. Data Test

Satisfaction level	last evaluation	Number Project	average montly hours	Time spend company	Work accident	left	promotion last 5years	job description	salary
0,76	0,73	3	158	2	0	0	0	Technical	

Langkah – langkah perhitungan menggunakan algoritma k-Nearest Neighbor (k-NN) yaitu sebagai berikut :

1. Menentukan Nilai K

Nilai K = 5

2. Menentukan jarak tetangga.

Untuk menentukan jarak tetangga rumus yang digunakan adalah rumus pada persamaan dua.

Berikut adalah langkah-langkah untuk menentukan jarak tetangga.

- a. Data *train* – data *test* ($X_1 - X_2$)

Pada tahap ini penulis melakukan penghitungan data dengan cara mengurangkan data *training* dengan data baru (data test). Pada penelitian ini data *training* adalah X_1 , dapat dilihat dari Tabel 5. dan data *test* adalah X_2 , dapat dilihat pada Tabel 6. Hasil yang diperoleh dari pengurangan data tersebut dapat dilihat pada Tabel 7. Hasil Pengurangan $X_1 - X_2$.

Contoh perhitungan: $(0.38 - 0.76) + (53 - 0.73) (2 - 3) + (157 - 158) + (3 - 2)$
 $= -0.38 + -0.2 + -1 + -1 + 1$

Tabel 7. Hasil Pengurangan $X_1 - X_2$

No.	Satisfaction level	Last evaluation	Number project	Average montly hours	Time spend company
1	-0,38	-0,2	-1	-1	1
2	0,04	0,13	2	262	4
3	-0,65	0,15	4	114	2
4	0,72	0,14	2	65	3
5	-0,39	-0,21	-1	1	1
6	-0,18	0,01	1	57	1
7	0,06	-0,06	-1	44	1

8	-0,31	-0,04	2	35	1
9	0,02	0,09	2	89	1
10	-0,27	-0,13	0	56	0

- b. Mengkuadratkan hasil Data *train* - data *test* ($X_1 - X_2$)

Setelah data training dikurang dengan data test tahapan selanjutnya adalah mengkuadratkan hasil $X_1 - X_2$, hasil yang diperoleh dapat dilihat pada

$$\begin{aligned} \text{Contoh perhitungan: } & (-0.38)^2 + (-0.2)^2 + (-1)^2 + (-1)^2 + (1)^2 \\ & = 0.1444 + 0.04 + 1 + 1 + 1 \end{aligned}$$

Tabel 8. Hasil Perpangkatan

No.	Satisfaction level	Last evaluation	Number project	Average monthly hours	Time spend Company
1	0,1444	0,04	1	1	1
2	0,0016	0,0169	4	68644	16
3	0,4225	0,0225	16	12996	4
4	0,5184	0,0196	4	4225	9
5	0,1521	0,0441	1	1	1
6	0,0324	0,0001	1	3249	1
7	0,0036	0,0036	1	1936	1
8	0,0961	0,0016	4	1225	1
9	0,0004	0,0081	4	7921	1
10.	0,0729	0,0169	0	3136	0

- c. Menjumlahkan hasil pengkuadratan dan mengakarkan hasil penjumlahan

Pada tahap ini penulis menjumlahkan data yang dikuadratkan secara keseluruhan dan mengakarkan hasil penjumlahan data tersebut, adapun hasil dari pengkuadratan dan hasil pengakaran data tersebut dapat dilihat pada Tabel 9. Hasil Penjumlahan dan Pengakaran

Contoh penjumlahan: $0.1444 + 0.04 + 1 + 1 + 1$

$$=\sqrt{3,1844}$$

Tabel 9. Hasil Penjumlahan dan Pengakaran

Hasil Penjumlahan	Hasil Perpangkatan
3,1844	1,784489
68664,02	262,0382
13016,45	114,0896
4238,538	65,10406
3,1962	1,787792
3251,033	57,01783
1938,007	44,0228
1230,098	35,07275
7926,009	89,02813
3136,09	56,0008

3. Mengurutkan ranking

Setelah jarak ditemukan pada tahap penghitungan jarak, langkah selanjutnya adalah mengurutkan rangking data secara *Ascending*, urutan rangking jarak dapat dilihat pada Tabel 10. Pengurutan Ranking.

Tabel 10. Pengurutan Ranking

No.	Hasil Penjumlahan (Σ)	Hasil Pengakaran ($\sqrt{}$)	Ranking
1	3,1844	1,784489	1
2	3,1962	1,787792	2
3	1230,098	35,07275	3
4	1938,007	44,0228	4
5	3136,09	56,0008	5
6	3251,033	57,01783	6
7	4238,538	65,10406	7
8	7926,009	89,02813	8
9	13016,45	114,0896	9
10	68664,02	262,0382	10

4. Mengambil nilai yang diranking sebanyak nilai K

Setelah melakukan tahap perengkingan data, tahap selanjutnya adalah mengambil data sebanyak jumlah k, dimana nilai data yang diambil adalah lima data terkecil. Data yang diambil sebanyak nilai k dapat dilihat pada Tabel 11. Urutan Lima Data

Tabel 11. Urutan Lima Data

No.	Satisfaction level	Last evaluation	Number project	Average Montly hours	Time Spend company	Satisfaction level	Last Evaluation	Numbe r project	average montly hours	time spend company	Jumlah (Σ)	Akar ($\sqrt{\quad}$)
1	-0,38	0,1444	-0,2	0,04	-1	1	-1	1	1	1	3,1844	1,784489
2	-0,39	0,1521	-0,21	0,0441	-1	1	1	1	1	1	3,1962	1,787792
3	-0,31	0,0961	-0,04	0,0016	2	4	35	1225	1	1	1230,098	35,07275
4	0,06	0,0036	-0,06	0,0036	-1	1	44	1936	1	1	1938,007	44,0228
5	-0,27	0,0729	-0,13	0,0169	0	0	56	3136	0	0	3136,09	56,0008

5. Menyesuaikan dengan *class*

Lima data yang telah diambil pada tahap pengambilan data sebanyak k disesuaikan dengan *class* masing-masing data pada data *training*. Hasil penyesuaian *class* dapat dilihat pada Tabel 12. Penyesuaian Class

Tabel 12. Penyesuaian Class

Satisfaction level	Last evaluation	Number project	Average Montly hours	Time Spend company	Left
-0,38	-0,2	-1	-1	1	1
-0,39	-0,21	-1	1	1	1
-0,31	-0,04	2	35	1	0
0,06	-0,06	-1	44	1	0
-0,27	-0,13	0	56	0	0

6. Kesimpulan

Dari lima nilai jarak yang diperoleh data menunjukkan dua karyawan keluar dan 3 karyawan tetap bertahan. Sehingga hasil dari prediksi yang dilakukan menggunakan algoritma k -NN menyimpulkan data karyawan tersebut adalah bertahan.

3.5 Analisis Algoritma Naive Bayes

Pada Tabel 13. Data Percobaan 1 Naive Bayes dan Tabel 14. Data Percobaan 2 Naive Bayes dapat dilihat contoh perhitungan manual dari data *turnover* karyawan menggunakan algoritma *naive bayes*.

Tabel 13. Data Percobaan 1 Naive Bayes

No	Satisfaction Level	Last evaluation	Number project	Average Monthly hours	Time Spend company	Work accident	Promotion Last 5 years	Job description	salary	left
1.	0.41	0.84	6	191	6	0	0	HR	medium	0
2.	0.41	0.55	2	148	3	0	0	Sales	low	1

$$\begin{aligned}
 \text{Likelihood(Left=0)} &= \frac{30}{60} * \frac{1}{30} * \frac{2}{30} * \frac{3}{30} * \frac{1}{30} * \frac{3}{30} * \frac{29}{30} * \frac{30}{30} * \frac{4}{30} * \frac{15}{30} \\
 &= 0.5 * 0.03 * 0.06 * 0.1 * 0.03 * 0.1 * 0.96 * 1 * 0.13 * 0.5 \\
 &= 0.000000016848
 \end{aligned}$$

$$\begin{aligned}
 \text{Likelihood(Left=1)} &= \frac{30}{60} * \frac{3}{30} * \frac{2}{30} * \frac{13}{30} * \frac{1}{30} * \frac{13}{30} * \frac{30}{30} * \frac{30}{30} * \frac{25}{30} * \frac{25}{30} \\
 &= 0.5 * 0.03 * 0.06 * 0.43 * 0.03 * 0.43 * 1 * 1 * 0.83 * 0.83 \\
 &= 0.000000011463
 \end{aligned}$$

$$P(\text{left}=0) = \frac{0.000000016848}{0.000000016848 + 0.00011463} = 0.001468$$

$$P(\text{left}=1) = \frac{0.00011463}{0.00001147} = 0.99$$

Hasil yang diperoleh dari perhitungan manual adalah $P(1) > P(0)$, jadi perhitungan tersebut diatas menunjukkan bahwa karyawan akan keluar dari perusahaan.

Tabel 14. Data Percobaan 2 Naive Bayes

No	Satisfaction Level	Last evaluation	Number project	Average Monthly hours	Time Spend company	Work accident	Promotion Last 5 years	Job description	salary	left
1.	0.82	0.67	2	202	3	0	0	Sales	low	0
2.	0.82	0.87	4	239	5	0	0	Sales	low	1

$$\begin{aligned}
 \text{Likelihood(Left=0)} &= \frac{30}{60} * \frac{1}{30} * \frac{1}{30} * \frac{5}{30} * \frac{1}{30} * \frac{17}{30} * \frac{30}{30} * \frac{30}{30} * \frac{14}{30} * \frac{15}{30} \\
 &= 0.5 * 0.03 * 0.03 * 0.16 * 0.03 * 0.56 * 1 * 1 * 0.46 * 0.5 \\
 &= 0.000000278208
 \end{aligned}$$

$$\begin{aligned}
 \text{Likelihood(Left=1)} &= \frac{30}{60} * \frac{1}{30} * \frac{2}{30} * \frac{3}{30} * \frac{1}{30} * \frac{7}{30} * \frac{30}{30} * \frac{30}{30} * \frac{27}{30} * \frac{25}{30} \\
 &= 0.5 * 0.03 * 0.06 * 0.1 * 0.03 * 0.23 * 1 * 1 * 0.9 * 0.83 \\
 &= 0.000000463887
 \end{aligned}$$

$$\begin{aligned}
 P(\text{left}=0) &= \frac{0.000000278208}{0.000000793638 + 0.000000278208} \\
 &= 0.025
 \end{aligned}$$

$$\begin{aligned}
 P(\text{left}=1) &= \frac{0.000000793638}{0.000001071846} \\
 &= 0.74
 \end{aligned}$$

Hasil yang diperoleh dari perhitungan manual pada percobaan dua adalah $P(1) > P(0)$, jadi perhitungan tersebut diatas menunjukkan bahwa karyawan akan keluar dari perusahaan.

3.6 Analisis Data Transformation

Data transformation digunakan pada metode *preprocessing*. Pada penelitian ini tipe data yang ditransformasikan adalah tipe data string menjadi tipe data integer atau label menjadi numerik.

Contoh data *transformation* pada *value salary*: *salary* memiliki 3 *value* yaitu *low*, *medium* dan *high*. Pada *value salary* data *transformation* dapat dilakukan dengan cara mengubah *salary low* menjadi 100, *salary medium* menjadi 010, dan *salary high* menjadi 001. Data *transformation* bertujuan agar proses data *mining* lebih efisien dan juga agar pola yang dihasilkan lebih mudah untuk dipahami, pada penelitian ini data *transformation* dilakukan agar simulator dapat mengolah nilai variabel yang ditransformasikan. Berikut adalah tabel hasil data transformasi.

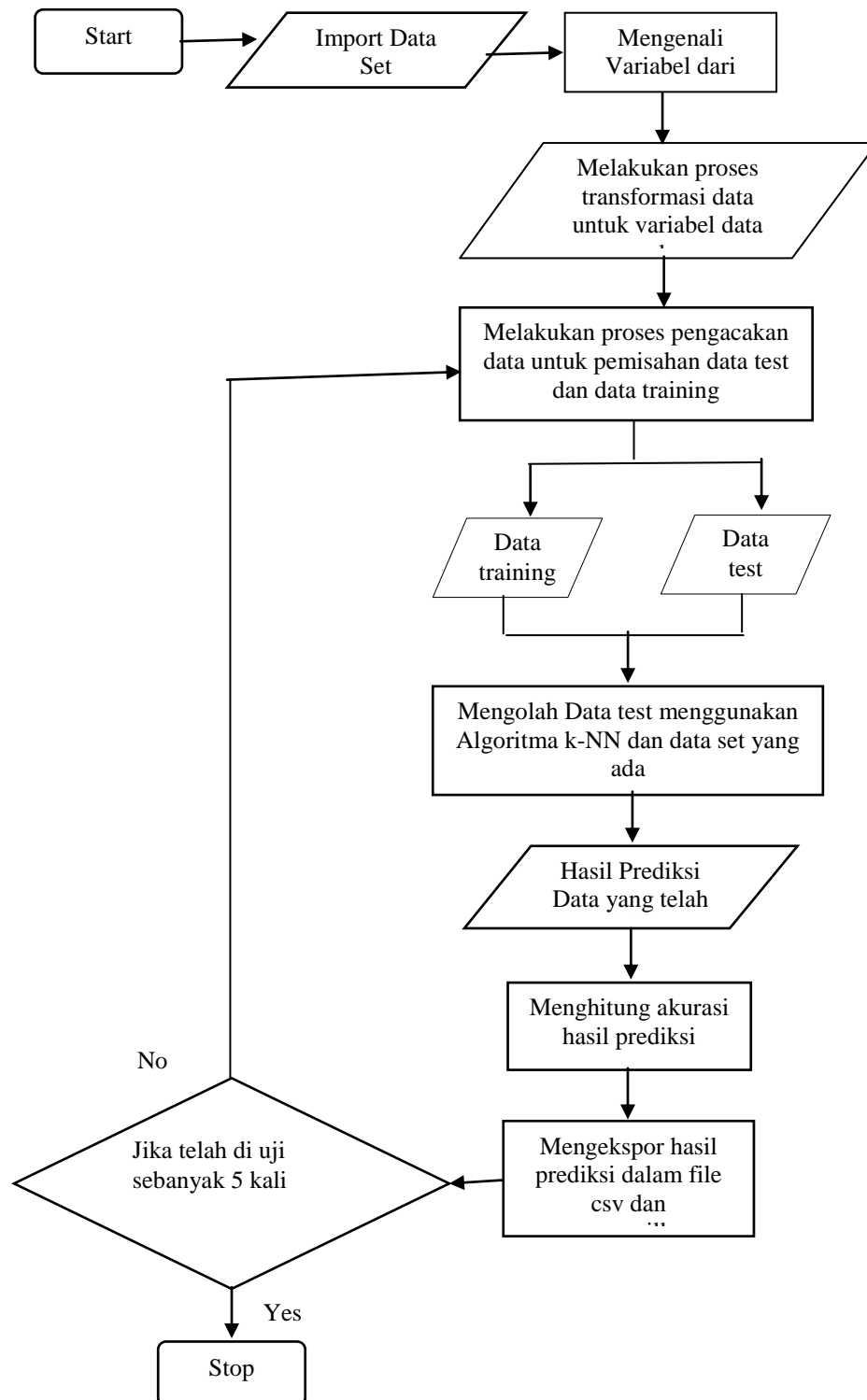
B	C	D
0	1	2
1	0	0
0	0	1
0	0	1
0	0	1
0	1	0
0	1	0
0	1	0
0	1	0
0	1	0
1	0	0

Gambar 3. Data Transformation

3.7 Desain Eksperimen dengan Algoritma KNN

Berikut adalah desain eksperimen menggunakan algoritma k-NN yang peneliti gunakan sebagai acuan dalam proses pengerjaan Tugas Akhir. Pada gambar desain berikut dijelaskan proses-proses yang akan dilakukan dalam pengerjaan

Tugas Akhir. Gambar 4 menjelaskan hal apa saja yang harus dilakukan selama tahap implementasi.

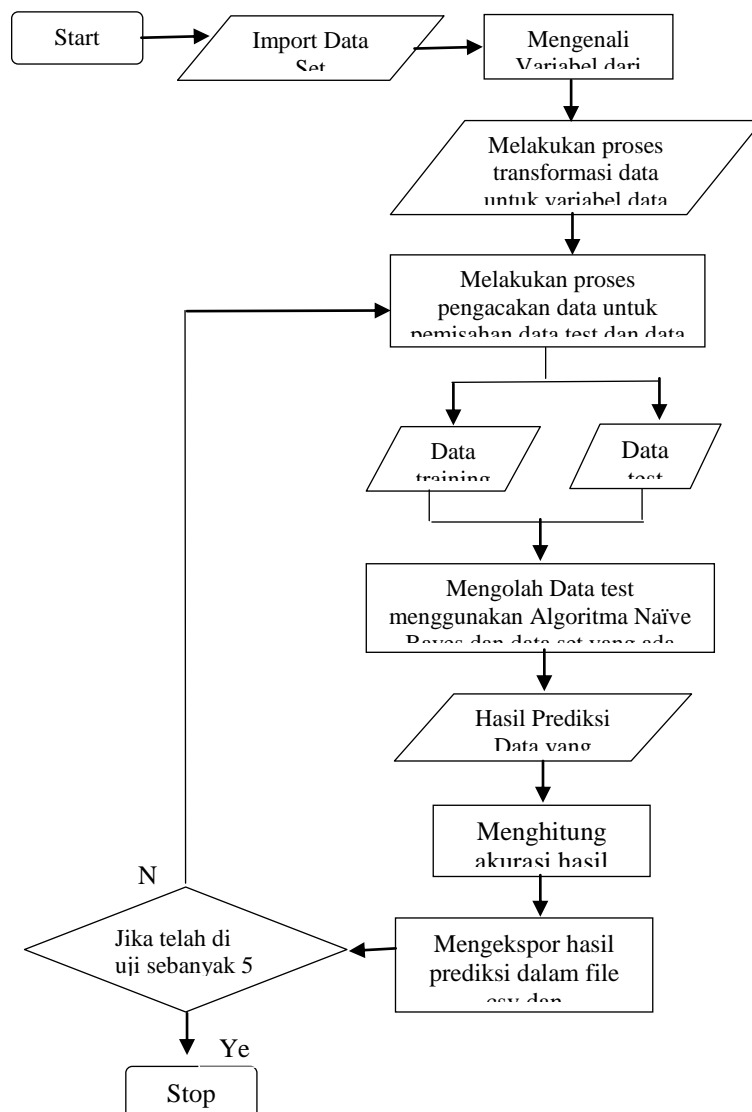


Gambar 4. Desain Eksperimen Algoritma KNN

Pada Gambar 4. dijelaskan proses eksperimen dari awal hingga akhir. Tahap awal pada proses eksperimen prediksi *turnover* adalah dengan mengimport data set yakni berupa data *turnover* untuk simulator, setelah data set diimport kemudian dilakukan proses input data uji dan mengolah data uji menggunakan algoritma k-NN dan kemudian menghasilkan prediksi data. Tahapan diatas akan dilakukan secara berulang untuk mendapatkan akurasi terbaik.

3.8 Desain Eksperimen dengan Algoritma Naïve Bayes

Berikut adalah desain eksperimen menggunakan algoritma *naive bayes* yang kami gunakan sebagai acuan dalam proses pengerjaan Tugas Akhir. Pada gambar desain berikut dijelaskan proses-proses yang akan dilakukan dalam pengerjaan Tugas Akhir.



Gambar 5. Desain Eksperimen dengan Algoritma Naïve Bayes

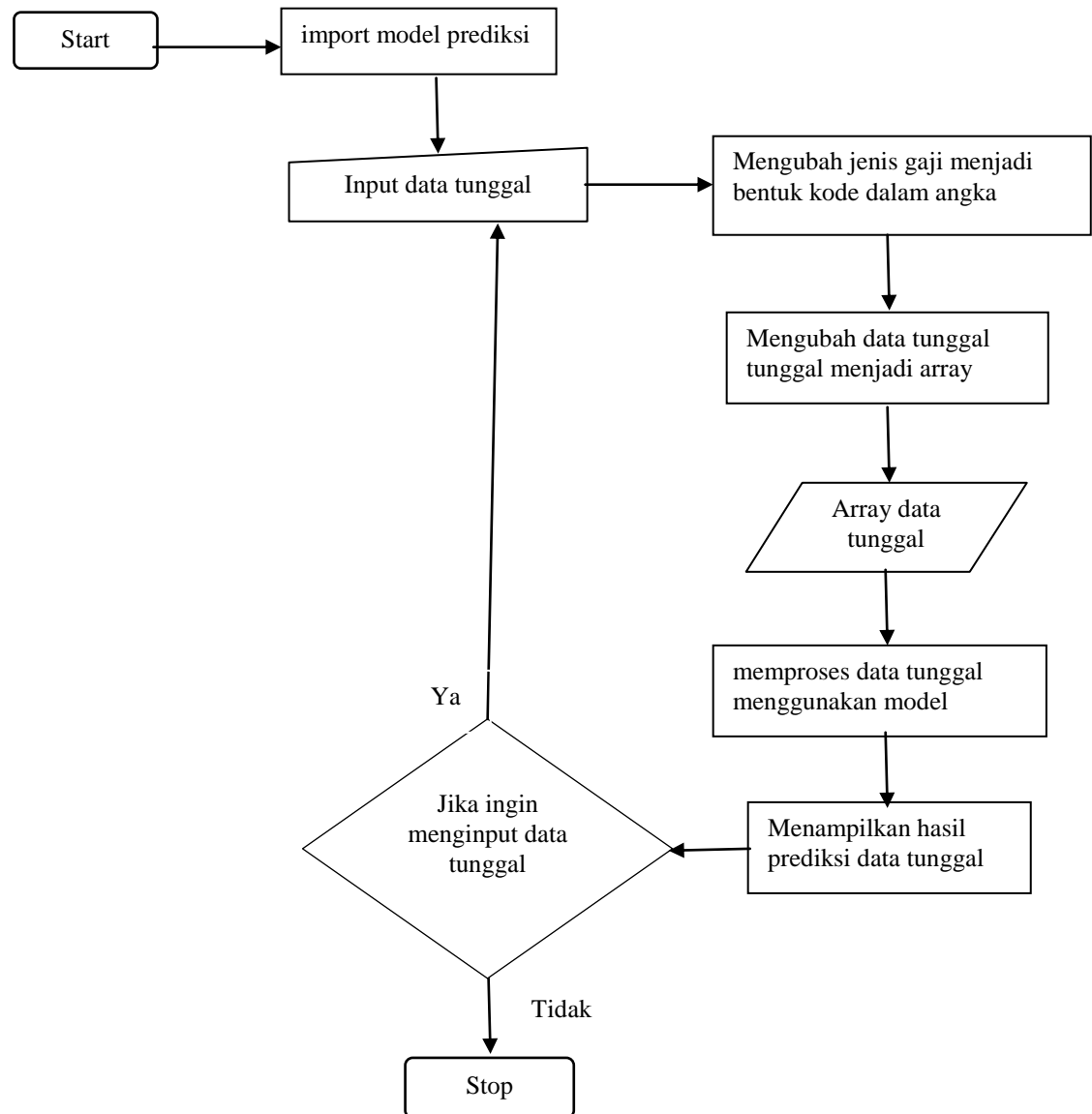
Pada Gambar 5. dijelaskan proses eksperimen menggunakan algoritma *naive bayes* dari awal hingga akhir. Tahap awal pada proses eksperimen prediksi *turnover* adalah dengan mengimport data set yakni berupa data *turnover* untuk simulator, setelah data set diimport tahapan selanjutnya adalah memisahkan *data train* dan *data test* kemudian *data train* akan diolah menggunakan algoritma *naive bayes*, kemudian menghasilkan prediksi data. Setelah prediksi berhasil ditampilkan tahapan selanjutnya adalah menampilkan akurasi hasil prediksi yang telah dilakukan pada tahap sebelumnya.

3.9 Desain Simulator

Design Interface merupakan gambaran simulator yang akan dibangun, Gambar 6 merupakan desain simulator yang digunakan untuk memprediksi hasil data karyawan yang diinput ke simulator dan akan memberikan hasil prediksi terhadap nilai yang telah di berikan ke simulator.

The image shows a software interface titled "Prediksi Turnover". On the left, under the heading "Variabel Nilai", there is a list of nine variables: Satisfaction Level, Last Evaluation, Number Projects, Average Monthly Hours, Time Spent Company, Work Accident, Promotion last 5 years, Job Description, and Sallary. Each variable has a corresponding empty text input box to its right. Below these inputs is a button labeled "Button". To the right of the input fields is a large rectangular box labeled "Hasil Prediksi Terhadap Nilai" with the subtitle "Hasil prediksi" inside it, intended for the simulation's output.

Gambar 6. Interface Simulator



Gambar 7. Data Flow Simulator

Data flow untuk simulator diatas menggambarkan cara kerja simulator. Setelah program di jalankan, hal yang pertama yang dilakukan oleh simulator adalah untuk meng-*import* model yang akan digunakan. Isi dari bagian model merupakan data set yang telah di training dan memiliki akurasi tinggi dan algoritma k-NN yang akan digunakan untuk memprediksi data tunggal yang akan di uji pada simulator.

Setelah data tunggal di input, simulator akan mengenali data yang telah di terima dan mengubah salah satu nilai dari data tunggal yaitu salary. Nilai salary yang sebelumnya merupakan dalam bentuk string akan dirubah menjadi bentuk kode, yaitu : 0,1,0 untuk nilai salary low, 0,0,1 untuk nilai salary medium, dan 1,0,0 untuk nilai salary high.

Sehingga data tunggal yang telah di kenali dan sebagian dari nilai data tunggal yang telah dirubah akan digabungkan kedalam sebuah array. Hal ini dilakukan agar nilai dapat proses oleh model prediksi yang telah dibuat di dalam simulator. Setelah simulator selesai memproses array data tunggal tersebut, simulator akan memberikan hasil prediksi dari data tunggal tersebut.

3.10 Use Case Scenario

Berikut adalah scenario penggunaan simulator prediksi *turnover* karyawan menggunakan algoritma knn.

Tabel 15. Use Case Scenario

<i>Use Case Description</i>	akan melakukan prediksi karyawan yang akan keluar atau bertahan diperusahaan dengan menginput nilai kedalam simulator yang telah menggunakan algoritma dengan nilai akurasi terbaik. <i>User</i> akan memasukkan satu per satu nilai variabel untuk melihat apakah karyawan akan keluar atau bertahan di perusahaan.	
<i>Actor</i>	<i>User</i>	
<i>Primary Flow of Events</i>	<i>Action</i>	<i>System Response</i>
	1. <i>User</i> memasukkan nilai variabel ke simulator.	
		2. Sistem menampilkan simulator
	3. <i>User</i> menginput nilai tiap variabel yang dibutuhkan oleh simulator untuk di prediksi	
	4. <i>User</i> menekan button prediksi	
		5. simulator memproses nilai variabel yang telah di input oleh <i>User</i>
		6. simulator menampilkan hasil prediksi terhadap nilai kepada <i>User</i>

BAB 4

ANALISIS DAN DESAIN

Pada bab ini dijelaskan mengenai implementasi dari program yang digunakan untuk melakukan prediksi *turnover* karyawan dengan menggunakan algoritma k-NN dan algoritma *Naive Bayes*. Pada bagian ini akan dipaparkan lingkungan operasional, pelaksanaan implementasi, hasil implementasi dan pembahasan.

4.1 Implementasi

Sub bab implementasi menjelaskan tahap implementasi dalam menghasilkan sistem aplikasi rekomendasi. Subbab ini mencakup lingkungan implementasi, batasan implementasi dan hasil implementasi.

4.1.1 Skenario Implementasi

Pada bagian ini akan dijelaskan mengenai tahapan implementasi yang dilakukan peneliti.

1. Mendapatkan data karyawan sebanyak 14.999.
2. Membagi data karyawan menjadi data tunggal dan data kolektif.
3. Melakukan *training* dan *testing* pada data kolektif
4. Melakukan pengujian pada data kolektif
5. Melakukan pengujian pada data tunggal
6. Mengitung akurasi pada pengujian data.
7. Menarik Kesimpulan.

4.1.2 Lingkungan Implementasi

Implementasi mencakup lingkungan perangkat keras dan perangkat lunak yang digunakan untuk membangun aplikasi. Spesifikasi perangkat yang digunakan pada pengimplementasi adalah sebagai berikut:

a. Hardware

Spesifikasi Hardware yang digunakan yaitu:

1. *Processor : Intel core i5*

2. *Memory* : 8.00 GB RAM

b. *Software*

1. Sistem Operasi : min Windows 8

2. IDE : *Spyder*

3. *Tools* : *Anocanda*

4. *Library* : Pandas, sklearn.preprocessing,
sklearn.cross_validation

5. *Browser* : UC Browser, Google Chrome, Mozilla Firefox

7. Bahasa Pemograman : Python

4.1.3 Batasan Implementasi

Pada Subbab ini menjelaskan batasan implementasi yang dilakukan. Batasan implementasi yang dilakukan adalah sebagai berikut:

1. Data yang digunakan untuk proses implementasi adalah 14.999 data karyawan.
2. Data yang digunakan selama tahap implementasi disimpan dalam format csv.

4.1.4 Pelaksanaan Implementasi

Pada subbab ini dijelaskan bagaimana pelaksanaan implementasi prediksi *turnover* karyawan. Pada saat *implementasi*, data karyawan akan melewati proses *training* dan proses *testing*. Data akan di-*training* dan di-*testing* satu persatu maupun secara keseluruhan. *Training* dan *testing* secara keseluruhan bertujuan untuk melakukan pengujian klasifikasi dan mengukur akurasi data. Adapun yang menjadi tujuan dari klasifikasi data dan pengukuran akurasi data bertujuan untuk menarik kesimpulan dari pengujian tersebut.

Langkah-langkah yang digunakan untuk melakukan prediksi *turnover* karyawan adalah:

1. Program dapat membaca file CSV
2. Program memisahkan data training dengan data test dalam kasus ini presentasi pengambilan data yang digunakan adalah 80% - 20%. 80 % digunakan untuk data training dan 20 % digunakan untuk data test.
3. Tahapan selanjutnya adalah program mempelajari data menggunakan model yang sudah ada.
4. Tahapan selanjutnya adalah melakukan pengujian terhadap data training.
5. Output dari pengujian adalah berupa hasil prediksi yang ditampilkan dalam file csv.

4.1.4 Proses Prediksi Pada Algoritma KNN

Berikut pemaparan daftar proses prediksi menggunakan algoritma k-NN yang dilakukan pada saat implementasi.

1. Import *data set*

Import *data set* dapat dilihat pada potongan kode pada Gambar 8. Potongan

```
dataset = pd.read_csv('E:/Kuliah/SEMVI/TA 2/ANALISIS/For_Sent/raw_data.csv')
X = dataset.iloc[:, [0,1,2,3,4,5,7,9]].values
y = dataset.iloc[:, 6].values
```

Kode Untuk Import Data Set

Gambar 8. Potongan Kode Untuk Import Data Set

Pada potongan kode Gambar 8. Potongan Kode Untuk Import Data Set dijelaskan data yang di import merupakan data dengan format CSV.

2. Pemisahan data *training* dan data *test*.

Potongan kode pemisahan data *training* dan data *test* dapat dilihat pada Gambar 9. Potongan Kode untuk Pemisahan Data Training dan Data Set

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=20)
```

Gambar 9. Potongan Kode untuk Pemisahan Data Training dan Data Set

Potongan kode pada Gambar 9. Potongan Kode untuk Pemisahan Data Training dan Data Set berfungsi untuk memisahkan data *training* dan data *test*. Pada potongan kode di atas dijelaskan data *test* yang ambil adalah 20 %, sisanya dimasukkan dalam data *training* yaitu 80 %.

3. Proses klasifikasi

Proses klasifikasi pada algoritma k-NN dapat dilihat pada Gambar 10. Potongan Kode untuk Klasifikasi.

```
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors = 75, metric = 'minkowski', p = 2)
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)
```

Gambar 10. Potongan Kode untuk Klasifikasi

Pada proses klasifikasi data diambil dengan algoritma K Neighbors Classifier. Pada penelitian ini nilai K yang diambil adalah 75. Untuk mencari data yang 75 digunakan *library minkowski*.

4. Melakukan pengecekan

Setelah data diklasifikasi, tahapan selanjutnya adalah melakukan pengecekan terhadap data yang telah diklasifikasi, proses pengecekan dapat dilihat pada potongan Gambar 11. Potongan Kode untuk Pengecekan

```
def run_cv(X_test,y_pred,clf_class, method, **kwargs):
    from sklearn.model_selection import cross_val_predict
    # Initialize a classifier with key word arguments
    clf = clf_class(**kwargs)
    predicted = cross_val_predict(clf, X_test, y_pred, cv=3, method=method)
    return predicted
```

Gambar 11. Potongan Kode untuk Pengecekan

Potongan kode pada Gambar 11. Potongan Kode untuk Pengecekan akan menghasilkan *output* seperti pada Gambar 12. Output Pengecekan

Name	Type	Size	Value
X	float64	(14999, 10)	Min: 0.0 Max: 310.0
X_test	float64	(3000, 10)	Min: 0.0 Max: 310.0
X_train	float64	(11999, 10)	Min: 0.0 Max: 310.0
data	DataFrame	(3000, 10)	Column names: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
dataset	DataFrame	(14999, 10)	Column names: satisfaction_level, last_evaluation, number_projects, av ...
y	int64	(14999,)	Min: 0 Max: 1
y_pred	int64	(3000,)	Min: 0 Max: 1
y_test	int64	(3000,)	Min: 0 Max: 1
y_train	int64	(11999,)	Min: 0 Max: 1

Gambar 12. Output Pengecekan

5. Penghitungan akurasi

Perhitungan akurasi dapat dilihat pada potongan kode pada Gambar 13. Potongan Kode Penghitungan Akurasi

```
def accuracy(y_pred, predicted):
    # NumPy interprets True and False as 1. and 0.
    return metrics.accuracy_score(y_pred, predicted)

print ("%3f" % accuracy(y_pred, run_cv(X_test,y_pred,KNN, method='predict')))
```

Gambar 13. Potongan Kode Penghitungan Akurasi

Output dari potongan kode pada Gambar 13. Potongan Kode Penghitungan Akurasi dapat dilihat pada gambar Gambar 14. Output Penghitungan Akurasi

```
In [42]: print ("%3f" % accuracy(y_pred,
...: run_cv(X_test,y_pred,KNN, method='predict'))
...:
0.962
```

Gambar 14. Output Penghitungan Akurasi

Pada implementasi penghitungan akurasi dengan algoritma k-NN, akurasi yang diperoleh adalah sekitar 95.8 %.

4.1.5 Proses Prediksi Pada Algoritma Naive Bayes

Berikut pemaparan daftar proses prediksi menggunakan algoritma Naive Bayes yang dilakukan pada saat implementasi.

1. Import *data set*

Import *data set* dapat dilihat pada potongan kode pada Gambar 15. Import Data Set dengan Naive Bayes

```
import pandas as pd
dataset = pd.read_csv('D:/KULIAH/Sem VI/TA2/DEsktop/New folder (2)/TA/TUTOR_ON_FIX_GOING/raw_data.csv')
X = dataset.iloc[:, [0,1,2,3,4,5,7,9]].values
y = dataset.iloc[:, 6].values
```

Gambar 15. Import Data Set dengan Naive Bayes

Pada potongan kode diatas dijelaskan data yang di import merupakan data dengan format CSV.

2. Pemisahan data *training* dan data *test*.

Potongan kode pemisahan data *training* dan data *test* dapat dilihat pada Gambar 16. Pemisahan data Training dengan Data Test dengan Naive Bayes

```
from sklearn.cross_validation import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 20)
```

Gambar 16. Pemisahan data Training dengan Data Test dengan Naive Bayes

Potongan kode di atas berfungsi untuk memisahkan data *training* dan data *test*. Pada potongan kode di atas dijelaskan data *test* yang ambil adalah 20 %, sisanya dimasukkan dalam data *training* yaitu 80 %.

3. Proses klasifikasi

Proses klasifikasi pada algoritma k-NN dapat dilihat pada potongan kode pada Gambar 17. Proses Klasifikasi dengan Naive Baye

```
from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train, y_train)

# Predicting the Test set results
y_pred = classifier.predict(X_test)
```

Gambar 17. Proses Klasifikasi dengan Naive Baye.

4. Melakukan pengecekan

Setelah data diklasifikasi, tahapan selanjutnya adalah melakukan pengecekan terhadap data yang telah diklasifikasi, proses pengecekan dapat dilihat pada potongan kode pada Gambar 18. Potongan Kode untuk Pengecekan dengan Naive Bayes.

```
def run_cv(X_test,y_pred,clf_class, method, **kwargs):
    from sklearn.model_selection import cross_val_predict
    # Initialize a classifier with key word arguments
    clf = clf_class(**kwargs)
    predicted = cross_val_predict(clf, X_test, y_pred, cv=3, method=method)
    return predicted
```

Gambar 18. Potongan Kode untuk Pengecekan dengan Naive Bayes

Potongan pada Gambar 18. Potongan Kode untuk Pengecekan dengan Naive Bayes akan menghasilkan *output* seperti pada Gambar 19. Output Pengecekan dengan Naive Bayes

Name *	Type	Size	Value
X	float64	(14999, 10)	Min: 0.0 Max: 310.0
X_test	float64	(3000, 10)	Min: 0.0 Max: 310.0
X_train	float64	(11999, 10)	Min: 0.0 Max: 310.0
data	DataFrame	(3000, 10)	Column names: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
dataset	DataFrame	(14999, 10)	Column names: satisfaction_level, last_evaluation, number_projects, av ...
y	int64	(14999,)	Min: 0 Max: 1
y_pred	int64	(3000,)	Min: 0 Max: 1
y_test	int64	(3000,)	Min: 0 Max: 1
y_train	int64	(11999,)	Min: 0 Max: 1

Gambar 19. Output Pengecekan dengan Naive Bayes

5. Penghitungan akurasi

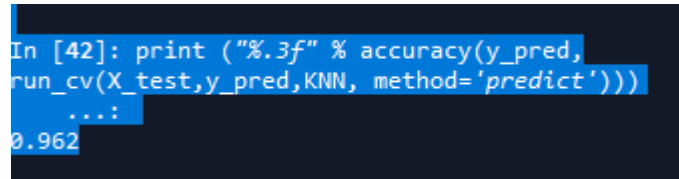
Perhitungan akurasi dapat dilihat pada potongan kode pada Gambar 19.

```
def accuracy(y_pred, predicted):
    # NumPy interprets True and False as 1. and 0.
    return metrics.accuracy_score(y_pred, predicted)

print ("%3f" % accuracy(y_pred, run_cv(X_test,y_pred,KNN, method='predict')))
```

Gambar 20. Penghitungan akurasi

Output dari potongan kode pada Gambar 19. dapat dilihat pada gambar Gambar 21. Output Penghitungan Akurasi



```
In [42]: print ('%.3f' % accuracy(y_pred,
run_cv(X_test, y_pred, KNN, method='predict')))
...:
0.962
```

Gambar 21. Output Penghitungan Akurasi

Pada implementasi penghitungan akurasi dengan algoritma k-NN, akurasi yang diperoleh adalah sekitar 95.8 %.

4.1.6 Proses Prediksi Data Tunggal

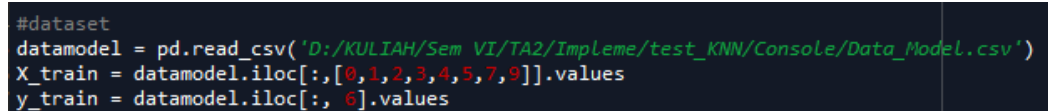
Pada subbab ini membahas tentang proses prediksi yang dilakukan untuk *single prediction* adapun proses implementasi yang dilakukan adalah sebagai berikut:

1. Menentukan data yang akan diprediksi

Pada proses ini diperlukan data yang akan diuji, data diambil dari data *test* yang telah dipisahkan.

2. Proses *import* data model

Proses *import* data model dapat dilihat pada potongan kode pada Gambar 22. Import Data Tunggal



```
#dataset
datamodel = pd.read_csv('D:/KULIAH/Sem VI/TA2/Impleme/test_KNN/Console/Data_Model.csv')
X_train = datamodel.iloc[:, [0,1,2,3,4,5,7,9]].values
y_train = datamodel.iloc[:, 6].values
```

Gambar 22. Import Data Tunggal

3. Proses *preprocessing* nilai *salary* dari data set

Proses *preprocessing* dapat dilihat pada potongan kode pada Gambar 23. Pengenalan Nilai Salary

```
#untuk mengenali nilai salary
labelencoder = LabelEncoder()
X_train[:,7]= labelencoder.fit_transform(X_train[:, 7])
onehotencoder = OneHotEncoder(categorical_features=[7])
X_train = onehotencoder.fit_transform(X_train).toarray()
```

Gambar 23. Pengenalan Nilai Salary

Setelah nilai *salary* dikenali tahapan selanjutnya adalah mengubah nilai *salary* seperti pada Gambar 25. Potongan Kode Input Nilai

```
X_value_1 = []
if(salary == 'low'):
    salary = [0,1,0]
elif(salary == 'medium'):
    salary = [0,0,1]
else:
    salary = [1,0,0]
```

Gambar 24. Potongan Kode Mengubah Nilai Salary

4. Proses input nilai

Proses input nilai dapat dilihat pada potongan kode pada Gambar 25. Potongan Kode Input Nilai

```
salary = input('salary : ')
satisfaction = input('satisfaction : ')
last = input('last : ')
number = input('number : ')
average = input('average : ')
time = input('time : ')
work = input('work : ')
promotion = input('promotion : ')
```

Gambar 25. Potongan Kode Input Nilai

Pada tahap ini nilai diinput secara manual adapun contohnya dapat dilihat pada Gambar 26. Contoh Input Nilai

```

salary : medium
satisfaction : 0.9
last : 09
number : 3
average : 123
time : 2
work : 0
promotion : 0
y_value : 1

```

Gambar 26. Contoh Input Nilai

5. Proses mengenali model data yang diinput

Proses pengenalan model dari data yang diinput dapat dilihat pada Gambar 27. Potongan Kode Pengenalan Model Data

```

clf=neighbors.KNeighborsClassifier(n_neighbors,weights='distance')
clf.fit(X_train, y_train)

```

Gambar 27. Potongan Kode Pengenalan Model Data

Pada proses ini program akan melihat variabel target dan variabel nilai yang diinput.

6. Proses prediksi tunggal

Proses prediksi dapat dilihat pada Gambar 28. Potongan Kode Prediksi Tunggal

```

print("Prediction Result :\n",y_predict)
if(y_predict==1):
    print("Karyawan keluar")
else:
    print("karyawn bertahan")
print("accruracy of prediction: ", scored*100,"%")

```

Gambar 28. Potongan Kode Prediksi Tunggal

Untuk hasil prediksi dapat dilihat pada Gambar 29. Hasil Prediksi

```

Prediction Result :
[1]
Karyawan keluar
accruracy of prediction: 100.0 %

```

Gambar 29. Hasil Prediksi

BAB 5

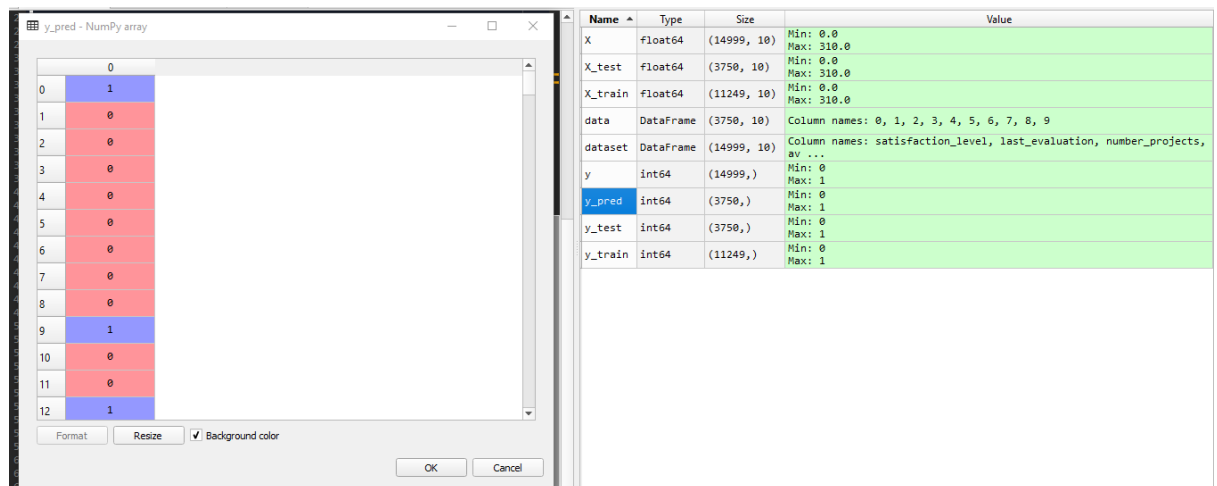
HASIL IMPEMETASI DAN PEMBAHASAN

5.1 Hasil Implementasi dan Pembahasan

Pada subbab ini membahas pembahasan dari keseluruhan hasil implementasi untuk data kolektif dan data tunggal yang dilakukan peneliti. Pada bagian ini peneliti akan membuat persentase akurasi prediksi *turnover* karyawan berdasarkan kategori data yang telah diimplementasikan.

5.2 Hasil Implementasi Data Kolektif dengan Algoritma K-Nearst Neighbor

Pada bagian ini dijelaskan mengenai hasil implementasi pada data kolektif dengan algoritma k-NN. Data *train* yang digunakan adalah 11.999 data karyawan dan data *test* 2.999. Hasil implementasi dapat dilihat pada **Error! Reference source not found..**



Gambar 30. Hasil Implementasi k-NN

Pada **Error! Reference source not found.** terdapat beberapa variabel diantaranya adalah variabel x, variabel y, variabel x_train, variabel y_train, variabel y_test dan y_pred. Variabel x berfungsi untuk mengenali parameter dan nilai parameter yang ada pada data set, sebelum dilakukan pemisahan antara data *training* dan data *test* sedangkan variabel y merupakan variabel target yang berfungsi untuk mengenali parameter dan nilai parameter target untuk diprediksi sementara variabel x_train merupakan variabel yang menyimpan data *train* yang digunakan pada proses prediksi dan variabel y_train merupakan variabel yang menyimpan nilai parameter *left* yang

akan dilatih untuk diprediksi. Hasil prediksi yang telah dilakukan akan disimpan pada variabel `y_pred`.

Pada implementasi algoritma k-NN diperoleh akurasi 95.8 % akurasi ini diperoleh dari jumlah seluruh akurasi pada tiap pengujian dibagi dengan jumlah pengujian. Hasil prediksi data dari tiap pengujian dapat dilihat pada Tabel 16. Hasil Prediksi dengan Algoritma KNN

Tabel 16. Hasil Prediksi dengan Algoritma KNN

No.	Pengujian	Akurasi (%)
1	I	95.8
2	II	95.5
3	III	95.7
4	IV	96.0
5	V	96.0
	Rata-rata	95.8

Secara detail hasil setiap pengujian adalah sebagai berikut:

1. Pengujian Pertama

Akurasi data diperoleh dari hasil pengujian pertama adalah 95.8 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 17. Komposisi Data Pengujian I KNN

Tabel 17. Komposisi Data Pengujian I KNN

Kategori	Data Set	Prediksi
Stay	2265	2136
Left	693	822
Akurasi		95.8 %

2. Pengujian Kedua

Akurasi data diperoleh dari hasil pengujian kedua adalah 95.5 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 18. Komposisi Data Pengujian II

Tabel 18. Komposisi Data Pengujian II KNN

Kategori	Data Set	Prediksi
----------	----------	----------

Stay	2265	2182
Left	693	776
Akurasi		95.5 %

3. Pengujian Ketiga

Akurasi data diperoleh dari hasil pengujian ketiga adalah 95.7 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 19. Komposisi Data Pengujian III KNN.

Tabel 19. Komposisi Data Pengujian III KNN

Kategori	Data Set	Prediksi
Stay	2250	2182
Left	708	776
Akurasi		95.7 %

4. Pengujian Keempat

Akurasi data diperoleh dari hasil pengujian keempat adalah 96.0 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada

Tabel 20. Komposisi Data Pengujian IV KNN

Kategori	Data Set	Prediksi
Stay	2254	2156
Left	704	802
Akurasi		96 %

5. Pengujian Kelima

Akurasi data diperoleh dari hasil pengujian kelima adalah 96.0 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 24. Komposisi Data Pengujian II Naive Bayes

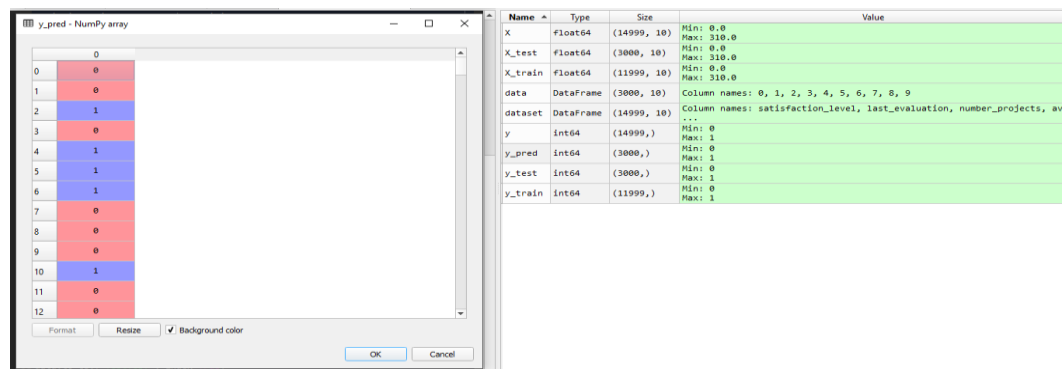
Tabel 21. Komposisi Data Pengujian V KNN

Kategori	Data Set	Prediksi
----------	----------	----------

Stay	2260	2187
Left	698	771
Akurasi		96.0%

5.4 Hasil Implementasi Data Kolektif dengan Algoritma Naive Bayes

Pada bagian ini dijelaskan mengenai hasil implementasi pada data kolektif dengan algoritma Naive Bayes. Data *train* yang digunakan adalah 11.831 data karyawan dan



data *test* 2.958.

Gambar 31. Hasil Implementasi Data Kolektif

Hasil implementasi dapat dilihat pada Gambar 30. dan hasil akurasi pada tiap pengujian dapat dilihat pada Tabel 22. Akurasi Tiap Pengujian dengan Algoritma Naive bayes Tabel 22. Akurasi Tiap Pengujian dengan Algoritma Naive bayes Pada implementasi algoritma Naive Bayes diperoleh akurasi 66,04 %, akurasi data diperoleh dari jumlah dari akurasi tiap pengujian dibagi jumlah banyak pengujian.

Tabel 22. Akurasi Tiap Pengujian dengan Algoritma Naive bayes

NO	Pengujian	Akurasi (%)
1	I	65.1
2	II	65.3
3	III	66.4
4	IV	66.8
5	V	66.6
	Rata-rata	66.04

Berikut adalah pembahasan dari hasil setiap pengujian:

1. Pengujian Pertama

Akurasi data diperoleh dari hasil pengujian pertama adalah 65.1 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 23. Komposisi Data Pengujian I Naive Bayes

Tabel 23. Komposisi Data Pengujian I Naive Bayes

Kategori	Data Set	Prediksi
Stay	2247	1263
Left	711	1695
Akurasi		65.1 %

2. Pengujian Kedua

Akurasi data diperoleh dari hasil pengujian kedua adalah 65.3 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 24. Komposisi Data Pengujian II Naive Bayes

Tabel 24. Komposisi Data Pengujian II Naive Bayes

Kategori	Data Set	Prediksi
Stay	2265	1691
Left	693	1267
Akurasi		65.3 %

3. Pengujian Ketiga

Akurasi data diperoleh dari hasil pengujian ketiga adalah 66.4 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 25. Komposisi Data Pengujian III Naive Bayes

Tabel 25. Komposisi Data Pengujian III Naive Bayes

Kategori	Data Set	Prediksi
Stay	2250	1683
Left	708	1275

Akurasi	5.3 %
---------	-------

4. Pengujian Keempat

Akurasi data diperoleh dari hasil pengujian keempat adalah 66.8 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 26. Komposisi Data Pengujian IV Naive Bayes

Tabel 26. Komposisi Data Pengujian IV Naive Bayes

Kategori	Data Set	Prediksi
Stay	2254	1673
Left	704	1285
Akurasi		66.8 %

5. Pengujian Kelima

Akurasi data diperoleh dari hasil pengujian kelima adalah 66.6 % pengujian ini diperoleh dari 2958 data karyawan secara detail komposisi data dapat dilihat pada Tabel 27. Komposisi Data Pengujian V Naive Bayes.

Tabel 27. Komposisi Data Pengujian V Naive Bayes

Kategori	Data Set	Prediksi
Stay	2260	1651
Left	698	1307
Akurasi		66.6%

5.4 Hasil Implementasi Data Tunggal

Pada bagian ini dijelaskan mengenai hasil implementasi pada data tunggal, jumlah data yang digunakan untuk implementasi data tunggal adalah 210 data. Data tunggal untuk simulator menggunakan algoritma dengan akurasi terbaik yaitu algoritma knn. Dari 210 data yang diuji diperoleh 12 data yang memiliki prediksi yang salah dengan data awal, enam diantaranya karyawan yang bertahan pada data awal dan enam sisanya adalah data karyawan yang keluar

pada data awal. Sementara 198 data memiliki prediksi yang benar, 198 data tersebut terdiri dari 161 data karyawan yang bertahan dan 37 data karyawan yang keluar.

Secara detail komposisi data dapat dilihat pada Tabel 28. Detail Data Tunggal

Tabel 28. Detail Data Tunggal

Kategori	Jumlah
Prediksi Benar	198
Prediksi Salah	12
Akurasi	94,28 %

Dari Tabel 28 diatas, diperoleh data akurasi prediksi adalah 94,28 % . Akurasi tersebut diperoleh dari perbandingan antara data prediksi yang benar dengan data prediksi yang salah, komposisi hasil prediksi dapat dilihat pada Tabel 29. Komposisi Data Tunggal.

Tabel 29. Komposisi Data Tunggal

No	Kategori Data	Prediksi	Jumlah
1	0 (Stay)	Benar	161
2	0 (Stay)	Salah	6
3	1 (Left)	Benar	37
4	1 (Left)	Salah	6

Berdasarkan Tabel 29 diatas diperoleh bahwa 161 data karyawan yang tinggal diperusahaan diprediksi benar dan 6 data karyawan adalah salah. Untuk data karyawan yang keluar dari perusahaan diprediksi 37 data adalah benar dan 6 data adalah salah.

BAB VI

SARAN DAN KESIMPULAN

Pada bab ini berisi kesimpulan dan saran yang diperoleh dari keseluruhan proses pengerjaan Tugas Akhir ini dan saran diharapkan dapat digunakan untuk pengembangan Tugas Akhir ini.

6.1 Kesimpulan

Dari hasil penelitian Tugas Akhir ini, dapat disimpulkan bahwa:

1. Simulator prediksi karyawan berhasil dibangun untuk dapat memprediksi karyawan yang keluar atau bertahan di perusahaan.
2. Algoritma k-NN dan naïve bayes diuji sebanyak lima kali untuk mengetahui tingkat akurasi masing masing algoritma. Algoritma k-NN menghasilkan akurasi 95.8 persen, sedangkan algoritma naïve bayes menghasilkan akurasi 66.04 persen. Algoritma terbaik berdasarkan proses pengujian adalah algoritma k-NN yaitu dengan akurasi 95.8 %.
3. Algoritma yang digunakan pada proses prediksi data tunggal adalah algoritma k-NN. Dengan prediksi data benar sebanyak 198 dengan akurasi 94.28 persen.

6.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan pada pengerjaan Tugas Akhir ini, maka saran yang perlu untuk melengkapi pengembangan selanjutnya adalah, pada penelitian ini, tingkat akurasi algoritma naïve bayes menghasilkan akurasi sebesar 66.04 persen. Algoritma naïve bayes memiliki kekurangan dalam proses kalkulasi data yang banyak, algoritma naïve bayes cenderung digunakan pada filtering kata. Disarankan untuk memilih algoritma yang mampu memiliki proses kalkulasi, misalnya *decision tree*.

References

- [1] W. Y. Z. H. CAI Juan, "Research on the Effects of EAPs on Turnover Intentions".
- [2] R. N. Muhammad Reza Anugrah Meilano, "Analisis Pengaruh Lingkungan Kerja Dan Kompensasi," vol. 6, pp. 1-11, 2017.
- [3] J. L. L. W. Y. B. Na AN, "Employee Satisfaction as an Important Tool in Human Resources Management," IEEE, 2008.
- [4] B. Cohen, "Human Resources Practices as Preictors of Work - Family Outcomes an Employee Turnover," 1997.
- [5] P. Joyner, "Python and Coding Theory," pp. 9-10, 2010.
- [6] J. Bell, "Machine Learning," in *Machine Learning*, Canada, John Wiley & Sons, Inc., 2015, p. 2.
- [7] H. I. B. A. F. Mutiara Ayu Banjarsari, "Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa," vol. 02, 2015.
- [8] A. Prawitasari, "Faktor-Faktor yang Mempengaruhi Turnover Intention Karyawan pada PT. Mandiri Tunas Finance Bengkulu".
- [9] R. S. S. e. al, "Employee turnover: a neural network solution," in *Computers & Operations Research* , USA, Science Direct, 2005, pp. 2635-2651.
- [10] X. D. Sumeet Dua, in *Data Mining and Machine Learning in Cybersecurity*, United States of America, CRC Press, 2011, p. 7.
- [11] Z. Wu, "Strategic Analysis of Employee Turnover," 2011.
- [12] C. G. Eddy M. Sutanto, "Kepuasan Kerja, Komitmen Organisasional dan," *Jurnal Mitra Ekonomi dan Manajemen Bisnis*, vol. 4, pp. 76-88, 2013.
- [13] H. Ongori, "A Review of The Literatur on Employee Turnover," *Business Management*, 2007.
- [14] A. Chandra, 12 3 2017. [Online]. Available: <https://www.datascience.or.id>. [Accessed 22 11 2017].
- [15] N. S. Yeni Kustiyahningsih, "Sistem Pendukung Keputusan Untuk Menentukan Jurusan Pada Siswa SMA Menggunakan Metode KNN dan SMART".

- [16] G. O. F. Mustakim, "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa," *Sains, Teknologi dan Industri*, 2016.
- [17] T. Chen, "Studies Computer Programing & Mathematics at National public school, HSR layout.," 09 10 2010. [Online]. Available: <https://www.quora.com/What-are-adventages-and-disadventages-of-Python>. [Accessed 15 11 2017].
- [18] J. Brownle, "A Tour of Machine Learning Algorithms," 25 11 2013. [Online]. Available: <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>. [Accessed 25 11 2011].
- [19] M. K. Jiawei Han, *Data Mining Concepts and Techniques*, San Francisco: Elsevier, 2006.
- [20] Samsuni, "Manajemen Sumber Daya Manusia," *Al Falah*, vol. XVIII, 2017.
- [21] S. B.-D. Shai Shalev-Shwartz, "Understanding Machine Learning:," in *Understanding Machine Leraning From Theory To Algorithms*, America, Cambridge, 2014, p. 347.
- [22] J. Brownlee, "A Tour of Machine Learning Algorithms," 25 November 2013. [Online]. Available: <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>. [Accessed 6 January 2018].
- [23] N. S. I. A. Arrie Kurniawan, "Klasifikasi Citra Batik Menggunakan Metode Ekstraksi Ciri yang Invarian Terhadap Rotasi," *Jurnal Ilmiah Teknologi Informasi*, vol. 12, 2014.
- [24] M. M. Deepak Kanojia, "Comparison of Naive Basian and K-NN Classifier," *International Journal of Computer Applications*, vol. 65, 2013.

Lampiran

No.	Satisfaction	Last Evaluation	Number Project	Average Monthly Hours	Time Spent Company	Work Accident	Promotion	Salary	Left	Prediction
1	0.9	0.74	3	193	3	0	0	medium	0	Benar
2	0.57	0.54	4	142	4	0	0	medium	0	Benar
3	0.95	0.77	5	199	10	1	0	medium	0	Benar
4	1	0.63	2	105	2	0	0	medium	0	Benar
5	0.19	0.48	6	178	3	1	0	medium	0	Benar
6	0.37	0.45	2	142	3	0	0	low	1	Benar
7	0.89	0.98	3	220	3	0	0	low	0	Benar
8	0.82	0.87	5	145	6	0	0	medium	0	Benar
9	0.52	0.96	4	170	2	0	0	low	0	Benar
10	0.62	0.64	3	165	3	0	0	high	0	Benar
11	0.65	0.76	4	193	2	0	0	medium	0	Benar
12	0.49	0.88	4	244	3	0	0	low	0	Benar
13	0.72	0.66	4	184	3	0	0	high	0	Benar
14	0.38	0.45	2	151	3	0	0	low	1	Benar
15	1	0.8	3	223	3	0	0	medium	0	Benar
16	0.98	0.58	4	133	3	0	0	low	0	Benar
17	0.65	0.6	3	148	2	0	0	low	0	Benar
18	0.72	0.7	3	163	3	1	0	medium	0	Benar
19	0.9	0.85	3	158	7	0	0	medium	0	Benar
20	0.68	0.92	4	209	6	0	0	low	0	Benar
21	0.57	0.99	3	205	3	0	0	low	0	Benar
22	0.67	0.5	4	173	2	1	0	high	0	Benar
23	0.76	0.9	4	263	5	0	0	low	1	Benar
24	0.68	0.67	3	228	2	0	0	medium	0	Benar
25	0.98	0.68	4	253	3	0	0	high	0	Benar
26	0.11	0.82	6	304	4	1	0	low	1	Benar
27	0.13	0.84	5	189	5	0	0	low	0	Benar
28	0.64	0.77	3	249	2	1	0	low	0	Benar
29	0.75	0.77	4	204	2	0	0	low	0	Benar

30	0.65	0.91	4	243	5	1	0	medium	0	Benar
31	0.44	0.51	2	140	3	0	0	medium	1	Benar
32	0.63	0.76	3	176	2	0	0	low	0	Benar
33	0.48	0.6	2	121	5	0	0	low	0	Benar
34	0.92	0.58	5	205	2	1	0	medium	0	Benar
35	0.75	0.53	3	154	2	0	0	medium	0	Benar
36	0.24	0.5	4	232	3	0	1	high	0	Benar
37	0.26	0.43	4	215	4	1	0	low	0	Benar
38	0.9	0.69	4	274	2	0	0	medium	0	Benar
39	0.69	0.93	3	228	4	0	0	medium	0	Benar
40	0.92	0.87	4	226	6	1	0	medium	1	Benar
41	0.39	0.53	2	127	3	0	0	medium	1	Benar
42	0.49	0.69	2	188	4	0	0	medium	0	Benar
43	0.36	0.46	2	132	3	0	0	low	1	Benar
44	0.36	0.61	4	166	4	0	0	low	0	Benar
No.	Satisfaction	Last Evaluation	Number Project	Average Monthly Hours	Time Spent Company	Work Accident	Promotion	Salary	Left	Prediction
45	0.72	1	4	169	3	0	0	medium	1	Salah
46	0.19	1	4	192	4	0	0	low	1	Salah
47	0.97	0.61	4	167	3	0	0	high	0	Benar
48	0.82	0.94	3	253	2	0	0	low	0	Benar
49	0.95	0.78	4	245	3	0	0	low	0	Benar
50	0.56	0.74	3	154	2	0	0	medium	0	Benar
51	0.92	0.67	2	252	2	0	0	low	0	Benar
52	0.75	0.94	5	217	2	0	0	medium	0	Benar
53	0.39	0.54	2	131	3	0	0	low	1	Benar
54	0.69	0.52	3	186	3	0	0	medium	0	Benar
55	0.66	0.66	4	255	5	0	0	low	0	Benar
56	0.76	0.65	4	195	3	0	0	low	0	Benar
57	0.97	0.94	3	180	3	0	0	low	0	Benar
58	0.14	0.62	4	158	4	1	0	low	1	Salah
59	0.75	0.85	3	250	6	0	0	medium	0	Benar
60	0.74	0.83	4	142	3	0	0	medium	0	Benar

61	0.31	0.87	6	240	3	1	0	low	0	Benar
62	0.45	0.55	2	148	3	0	0	low	1	Benar
63	0.71	0.62	3	258	2	0	0	medium	0	Benar
64	0.37	0.54	2	145	3	0	0	medium	1	Benar
65	0.44	0.48	2	143	3	0	0	medium	1	Benar
66	0.91	0.97	5	233	5	0	0	medium	1	Benar
67	0.6	0.61	2	275	2	0	0	low	0	Benar
68	0.94	0.62	4	191	3	0	0	low	0	Benar
69	0.68	0.69	3	185	2	0	0	low	0	Benar
70	0.97	0.66	5	234	6	0	0	low	0	Salah
71	0.77	0.57	5	162	3	0	0	high	0	Benar
72	0.96	0.71	4	210	3	0	0	medium	0	Benar
73	0.11	0.74	6	290	5	0	0	medium	1	Benar
74	0.58	0.76	3	220	3	0	0	medium	0	Benar
75	0.11	0.94	6	264	4	0	0	low	1	Benar
76	0.09	0.96	6	245	4	0	0	low	1	Benar
77	0.42	0.4	3	160	2	0	0	low	0	Salah
78	0.75	0.59	4	199	2	0	0	low	0	Benar
79	0.57	1	4	227	10	0	0	low	0	Salah
80	0.89	0.57	4	176	4	0	0	high	0	Benar
81	0.94	0.89	2	118	4	0	0	medium	0	Benar
82	0.5	0.75	5	178	6	0	0	medium	0	Benar
83	0.76	0.72	5	228	2	0	0	high	0	Benar
84	0.49	0.8	2	275	2	0	0	low	0	Benar
85	0.75	0.83	4	133	4	0	0	high	0	Benar
86	0.8	0.8	3	175	2	0	0	medium	0	Benar
87	0.82	0.97	3	137	3	0	0	high	0	Benar
88	0.34	0.92	3	206	3	0	0	medium	0	Benar
89	0.92	0.93	4	247	2	0	0	medium	0	Benar
90	1	0.67	5	241	4	0	0	medium	0	Benar
91	0.31	0.87	4	184	3	0	0	low	1	Salah
92	0.53	0.5	6	264	6	0	0	low	0	Salah
93	0.62	0.91	3	269	8	0	0	medium	0	Benar
94	0.93	0.63	4	143	3	1	0	low	0	Benar

No.	Satisfaction	Last Evaluation	Number Project	Average Monthly Hours	Time Spent Company	Work Accident	Promotion	Salary	Left	Prediction
95	0.96	0.8	4	145	2	0	0	medium	0	Benar
96	0.98	0.64	3	263	4	0	0	medium	0	Benar
97	0.55	0.82	4	134	6	0	0	medium	0	Benar
98	0.94	0.9	5	191	4	0	0	low	0	Benar
99	0.73	0.86	3	200	4	0	0	medium	0	Benar
100	0.15	0.87	4	194	4	0	0	low	0	Benar
101	0.27	0.76	4	108	10	0	0	medium	0	Benar
102	0.79	0.36	3	114	3	0	0	high	0	Benar
103	0.76	0.74	4	261	3	0	0	medium	0	Benar
104	0.96	0.37	3	111	2	0	0	high	0	Benar
105	0.37	0.49	2	153	3	0	0	low	1	Benar
106	0.71	0.63	3	204	3	0	0	low	0	Benar
107	0.69	0.98	3	168	2	0	0	medium	0	Benar
108	0.66	0.96	3	243	3	1	0	medium	0	Benar
109	0.46	0.49	2	148	3	0	0	low	1	Benar
110	0.88	0.7	4	146	4	0	0	low	0	Benar
111	0.55	0.64	4	201	10	1	0	medium	0	Benar
112	0.73	0.41	2	231	6	1	0	medium	0	Benar
113	0.52	0.95	5	234	3	0	0	low	0	Benar
114	0.6	0.91	5	211	3	0	0	low	0	Benar
115	0.48	0.49	3	268	3	0	0	low	0	Benar
116	0.74	0.79	5	121	5	0	1	high	0	Benar
117	0.51	0.51	4	189	3	0	0	low	0	Benar
118	0.62	0.49	4	175	3	0	0	medium	0	Benar
119	0.81	0.83	3	133	3	1	0	medium	0	Benar
120	0.8	0.64	5	204	2	0	0	low	0	Benar
121	0.8	0.67	3	178	3	0	0	low	0	Benar
122	0.63	0.5	4	167	3	1	0	medium	0	Benar
123	0.63	0.65	3	162	2	0	0	low	0	Benar
124	0.81	0.8	3	183	2	1	0	medium	0	Benar
125	0.63	0.74	4	155	2	0	0	medium	0	Benar

126	0.86	0.87	5	171	2	0	0	low	0	Benar
127	0.67	0.66	6	272	3	0	0	low	0	Benar
128	0.75	0.99	5	254	5	0	0	low	1	Benar
129	0.93	0.74	4	201	3	1	0	medium	0	Benar
130	0.21	0.37	4	129	4	0	0	medium	0	Salah
131	0.67	0.77	5	167	2	0	0	medium	0	Benar
132	0.66	0.77	4	188	2	0	0	low	0	Benar
133	0.79	0.91	5	271	5	0	0	low	1	Benar
134	0.86	0.68	4	170	2	0	0	low	0	Benar
135	0.76	0.77	4	169	10	0	0	medium	0	Benar
136	0.1	0.92	7	307	4	0	0	low	1	Benar
137	0.75	0.86	3	249	3	0	0	low	0	Benar
138	0.71	0.78	4	227	2	0	0	low	0	Benar
139	1	0.74	2	187	3	0	0	medium	0	Benar
140	0.76	0.5	3	258	3	0	0	low	0	Benar
141	0.67	0.54	2	141	2	0	0	low	0	Salah
142	0.49	0.63	4	213	3	0	0	Medium	0	Benar
143	0.54	0.84	4	223	3	0	0	Medium	0	Benar
144	0.81	0.9	4	270	10	0	0	Medium	0	Benar
145	0.41	0.54	2	152	3	0	0	Low	1	Benar
No.	Satisfaction	Last Evaluation	Number Project	Average Monthly Hours	Time Spent Company	Work Accident	Promotion	Salary	Left	Prediction
146	0.75	0.79	4	176	4	1	0	Medium	0	Benar
147	0.6	0.53	2	235	7	0	0	Medium	0	Benar
148	0.54	0.79	5	249	3	1	0	High	0	Benar
149	0.77	0.9	4	104	5	0	0	Low	0	Benar
150	0.64	0.53	3	216	3	0	0	High	0	Benar
151	0.78	0.81	4	232	3	0	0	Medium	0	Benar
152	0.43	0.57	2	135	3	0	0	Low	1	Benar
153	0.39	0.55	2	159	3	0	0	High	1	Benar
154	0.29	0.4	4	138	4	0	0	Low	0	Benar
155	0.88	0.83	4	273	10	0	0	Medium	0	Benar
156	0.62	0.66	3	208	3	0	0	Low	0	Benar

157	0.59	0.53	4	213	2	0	0	Low	0	Benar
158	0.39	0.47	2	131	3	0	0	Low	1	Benar
159	0.89	0.79	4	152	3	0	0	Low	0	Benar
160	0.49	0.6	3	191	3	1	0	High	0	Benar
161	0.76	0.57	4	266	3	0	0	Low	0	Benar
162	0.1	0.83	6	295	4	0	0	Medium	1	Benar
163	0.94	0.97	5	197	5	1	0	Low	0	Benar
164	0.91	0.98	4	146	3	0	0	Low	0	Benar
165	0.5	0.78	4	208	3	1	0	Medium	0	Benar
166	0.43	0.75	3	108	2	0	0	Low	0	Benar
167	0.78	0.57	3	177	3	0	0	Medium	0	Benar
168	0.72	0.8	3	158	3	0	0	Medium	0	Benar
169	0.65	1	4	249	3	0	0	Medium	0	Benar
170	0.74	0.92	4	261	5	0	0	Medium	1	Benar
171	0.24	0.55	6	231	4	0	0	Low	0	Benar
172	0.61	0.88	5	146	7	1	0	Low	0	Benar
173	0.76	1	5	219	5	0	0	Low	1	Benar
174	0.91	0.9	3	169	3	0	0	Low	0	Benar
175	0.81	0.82	4	167	2	0	0	Medium	0	Benar
176	0.38	0.74	3	159	4	0	0	Medium	0	Benar
177	0.57	0.91	4	252	4	0	0	Low	0	Benar
178	0.49	0.96	3	140	3	0	0	Medium	0	Benar
179	0.73	0.85	5	224	4	0	0	Low	0	Benar
180	0.39	0.64	3	152	2	0	0	Low	0	Benar
181	0.81	0.7	6	161	4	0	0	Medium	1	Salah
No.	Satisfaction	Last Evaluation	Number Project	Average Monthly Hours	Time Spent Company	Work Accident	Promotion	Salary	Left	Prediction
182	0.88	0.76	4	159	4	1	0	Low	0	Benar
183	0.48	0.89	4	145	3	0	0	Low	0	Benar
184	0.64	0.59	3	200	2	1	0	Medium	0	Benar
185	0.36	0.78	6	151	3	0	0	Low	0	Benar
186	0.63	0.79	4	228	4	0	0	Low	0	Benar
187	0.37	0.54	2	152	3	0	0	Medium	1	Benar

188	0.7	0.73	5	151	2	0	0	Low	0	Benar
189	0.94	0.48	4	218	3	0	0	Medium	0	Benar
190	0.8	0.62	3	191	3	0	0	Medium	0	Benar
191	0.86	0.95	4	270	5	0	0	Low	1	Benar
192	0.55	0.67	3	150	2	0	1	Low	0	Benar
193	0.79	1	4	229	5	1	0	Medium	1	Benar
194	0.84	0.53	3	210	4	1	0	Medium	0	Benar
195	0.8	0.75	3	268	2	0	0	Medium	1	Salah
196	0.89	0.51	4	185	3	1	0	High	0	Benar
197	0.52	0.76	4	106	2	1	0	Low	0	Benar
198	0.92	0.77	5	217	4	0	0	Low	0	Benar
199	0.44	0.52	2	137	3	0	0	Medium	1	Benar
200	0.6	0.54	3	265	3	1	0	Medium	0	Benar
201	0.68	0.51	4	224	2	0	0	Low	0	Benar
202	0.55	0.71	3	211	2	0	0	Low	0	Benar
203	0.43	0.76	6	154	2	0	0	High	0	Benar
204	0.38	0.59	4	276	2	0	0	Low	0	Benar
205	0.8	0.97	4	259	2	0	0	Low	0	Benar
206	0.75	1	5	223	6	0	0	Medium	1	Benar
207	0.92	0.93	5	240	5	0	0	Low	1	Benar
208	0.8	0.87	5	240	5	0	0	Low	1	Benar
209	0.42	0.53	2	131	3	0	0	Medium	1	Benar
210	0.77	0.9	4	237	5	0	0	Low	1	Benar