

**ANALISIS HASIL EVALUASI KINERJA KELOMPOK JABATAN
FUNGSIONAL PEMERIKSA PAJAK KANTOR PELAYANAN
PAJAK PMA DUA MENGGUNAKAN METODE *NAÏVE BAYES***

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:

Agustinus Ryan Wicaksono

NIM: 175150400111031



PROGRAM STUDI SISTEM INFORMASI
JURUSAN SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2021

PENGESAHAN

ANALISIS HASIL EVALUASI KINERJA KELOMPOK JABATAN FUNGSIONAL
PEMERIKSA PAJAK KANTOR PELAYANAN PAJAK PMA DUA MENGGUNAKAN
METODE *NAÏVE BAYES*

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh:
Agustinus Ryan Wicaksono
NIM: 175150400111031

Skripsi ini telah diuji dan dinyatakan lulus pada
16 Juli 2021

Telah diperiksa dan disetujui oleh:

Dosen Pembimbing 1



Satrio Hadi Wijoyo, S.Si., S.Pd., M.Kom.
NIK: 2016098909101001

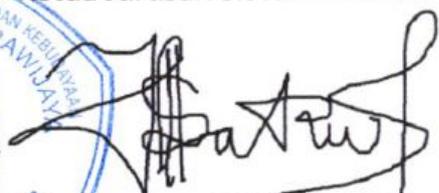
Dosen Pembimbing 2



Nanang Yudi Setiawan, S.T., M.Kom.
NIP: 197606192006041001

Mengetahui

Ketua Jurusan Sistem Informasi



Issa Arwani, S.Kom, M.Sc.
NIP: 198309222012121003

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar referensi.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 22 Juli 2021



Agustinus Ryan Wicaksono
NIM: 175150400111031

ABSTRAK

Agustinus Ryan Wicaksono, Analisis Hasil Evaluasi Kinerja Kelompok Jabatan Fungsional Pemeriksa Pajak Kantor Pelayanan Pajak PMA Dua Menggunakan Metode Naïve Bayes

Pembimbing: Satrio Hadi Wijoyo, S.Si., S.Pd., M.Kom. dan Nanang Yudi Setiawan, S.T., M.Kom.

Evaluasi penilaian kinerja merupakan proses evaluasi yang dilakukan oleh sebuah instansi baik instansi pemerintah maupun instansi swasta untuk mengukur performa serta kualitas pegawai di instansi tersebut. Kantor Pelayanan Pajak PMA Dua merupakan salah satu instansi pemerintah yang menjalankan evaluasi penilaian kinerja untuk menilai kinerja Fungsional Pemeriksa Pajak. Namun dalam pelaksanaan evaluasi penilaian kinerja, KPP PMA Dua menemui hambatan dimana banyaknya variabel data yang digunakan serta masih banyak data yang tersebar. Sehingga perlu ketelitian dalam mengolah data yang digunakan untuk menilai kinerja karyawan. Oleh karena itu penulis melakukan analisis menggunakan *naïve bayes* pada data penilaian kinerja fungsional pemeriksa pajak agar dapat mengetahui bagaimana kinerja pegawai tersebut. Penulis menggunakan data pegawai tahun 2019 sebanyak 64 data yang dimanfaatkan untuk melatih model *naïve bayes*. Hasil dari penerapan *data mining* menghasilkan akurasi sebesar 93,75% dengan menggunakan metode *Accuracy Score* dan akurasi sebesar 82,5% dengan menggunakan metode *K-Fold Cross Validation*. Model *naïve bayes* yang telah dibangun juga dimanfaatkan untuk memprediksi kinerja karyawan dengan menggunakan data pegawai tahun 2020 sebanyak 52 data. *Dashboard* yang telah dibentuk mendapatkan nilai rata-rata sebesar 80 yang berarti *dashboard* tersebut dapat diterima dan dapat digunakan oleh Kantor Pelayanan Pajak PMA Dua.

Kata kunci: Evaluasi Penilaian Kinerja, *Data Mining*, Klasifikasi, *Naïve Bayes*, *Cross Validation*, *Dashboard*.

ABSTRACT

Agustinus Ryan Wicaksono, *Analysis of Performance Evaluation Results of Functional Position Group Tax Audit at PMA Dua Tax Service Office using the Naïve Bayes Method*

Supervisors: Satrio Hadi Wijoyo, S.Si., S.Pd., M.Kom. and Nanang Yudi Setiawan, S.T., M.Kom.

Employee Performance Evaluation is an evaluation process conducted by an agency, both government agencies and private agencies to measure the performance and quality of employees in the agency. PMA Dua Tax Service Office is one of the government agency that run the performance evaluation to measure the performance of the Tax Auditor Functional. However, in the implementation of the performance evaluation, PMA Dua Tax Service Office encountered obstacles where the data variables were used and there were still a lot of scattered data. So it needs accuracy in processing the data used to assess employee performance. Therefore, the author conducted an analysis using naive bayes on the tax audit functional performance evaluation data in order to fund out how the employee's performance was. The author uses the 2019 employee data as much as 64 data which is used to train the naïve bayes model. The resultsof the application of data mining produce an accuracy of 93.75% using the Accuracy Score method and an accuracy of 82,5% using the K-Fold Cross Validation method. The naïve bayes model that has been built is also used to predict employee performance in 2020 using 52 data. The dashboard that has been formed has an average score of 80, which means that the dashboard is acceptable and can be used by the PMA Dua Tax Service Office.

Keywords: *Performance Evaluation, Data Mining, Classification, Naïve Bayes, Cross Validation, Dashboard.*

DAFTAR ISI

PENGESAHAN	ii
PERNYATAAN ORISINALITAS.....	iii
PRAKATA.....	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR KODE PROGRAM.....	xii
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Batasan Masalah	3
1.6 Sistematika Pembahasan.....	4
BAB 2 LANDASAN KEPUSTAKAAN.....	5
2.1 Kajian Pustaka	5
2.2 Profil KPP PMA Dua	5
2.2.1 Visi.....	6
2.2.2 Misi.....	6
2.2.3 Struktur Organisasi	7
2.3 <i>Data Mining</i>	7
2.3.1 Tahapan Data Mining.....	7
2.3.2 Metode Data Mining.....	8
2.4 Klasifikasi	9
2.5 <i>Naïve Bayes</i>	10
2.6 <i>K-Fold Cross Validation</i>	10
2.7 <i>Imbalanced Dataset</i>	11
2.8 <i>Dashboard</i>	11

2.9 <i>System Usability Scale (SUS)</i>	12
BAB 3 METODOLOGI	14
3.1 Identifikasi Masalah	15
3.2 Studi Literatur	15
3.3 Pengumpulan Data	15
3.4 Pengolahan Data	15
3.5 Pengujian dan Analisis Hasil	16
3.6 Visualisasi Hasil	16
3.7 Kesimpulan dan Saran	16
BAB 4 Implementasi.....	17
4.1 Pengumpulan Data.....	17
4.2 <i>Data Preprocessing</i>	19
4.2.1 <i>Import Data</i>	20
4.2.2 <i>Data Selection & Data Cleaning</i>	21
4.3 Penerapan <i>Naïve Bayes</i>	24
4.4 Hasil Akurasi Klasifikasi <i>Naïve Bayes</i>	25
4.5 Prediksi Data	26
4.6 <i>Dashboard</i> Kinerja Pegawai.....	28
BAB 5 HASIL DAN PEMBAHASAN	29
5.1 Hasil Uji Validitas.....	29
5.1.1 <i>Accuracy Score</i>	29
5.1.2 <i>K-Fold Cross Validation</i>	30
5.2 Prediksi Nilai Pegawai.....	30
5.3 Hasil Visualisasi Dashboard	32
5.4 Hasil <i>System Usability Scale</i>	34
BAB 6 PENUTUP	35
6.1 Kesimpulan	35
6.2 Saran.....	35
DAFTAR REFERENSI	36
LAMPIRAN A HASIL WAWANCARA	38
LAMPIRAN B HASIL <i>System usability scale</i>	39

DAFTAR GAMBAR

Gambar 2.1 Struktur Organisasi KPP PMA Dua	7
Gambar 2.2 Tahapan Data Mining.....	8
Gambar 2.4 Konsep K-Fold Cross Validation.....	11
Gambar 2.5 Tampilan Google Data Studio.....	12
Gambar 2.6 Skala Penilaian SUS.....	13
Gambar 3.1 Diagram Tahapan Metodologi Penelitian	14
Gambar 4.1 Data Fungsional Pemeriksa Pajak Tahun 2019.....	17
Gambar 4.2 Data Penilaian Tahun 2019	20
Gambar 4.3 Struktur Dataset Penilaian Sebelum Data Selection	23
Gambar 4.4 Struktur Dataset Penilaian Setelah Data Selection	23
Gambar 4.5 Struktur Dataset Penilaian Hasil Data Selection	23
Gambar 5.1 Hasil Prediksi Model Naïve Bayes.....	29
Gambar 5.2 Nilai Variable y_{test}	29
Gambar 5.3 Hasil Akurasi Menggunakan Accuracy Score	30
Gambar 5.4 Hasil Akurasi Menggunakan K-Fold Cross Validation	30
Gambar 5.5 Dashboard Penilaian Kinerja Pegawai Tahun 2019	32
Gambar 5.6 Dashboard Penilaian Kinerja Pegawai Tahun 2020	33
Gambar 5.7 Dashboard Hasil Prediksi Kinerja Pegawai Tahun 2020.....	33

DAFTAR TABEL

Tabel 2.1 Pertanyaan <i>System Usability Scale</i>	12
Tabel 4.1 Keseluruhan Atribut Dataset	18
Tabel 4.2 Atribut Dataset yang digunakan	19
Tabel 4.3 Data Hasil <i>Preprocessing</i>	24
Tabel 5.1 Data Hasil Prediksi	31
Tabel 5.2 Data Hasil Pengisian <i>System Usability Scale</i>	34
Tabel 5.3 Hasil Perhitungan <i>System Usability Scale</i>	34

DAFTAR KODE PROGRAM

Kode Program 4.1 <i>Import Data</i>	20
Kode Program 4.2 Data Selection & Data Cleaning.....	22
Kode Program 4.3 Penerapan Naïve Bayes.....	25
Kode Program 4.4 Hasil Akurasi <i>Naïve Bayes</i>	26
Kode Program 4.5 Prediksi Data Tahun 2020.....	28
Kode Program 4.6 <i>Export Data</i>	28

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Evaluasi penilaian kinerja pegawai merupakan sebuah kegiatan yang dilakukan untuk menilai dan mengevaluasi kemampuan pegawai serta melihat perkembangan pegawai itu sendiri. Selain itu juga evaluasi penilaian kinerja pegawai dilakukan untuk melihat apakah pegawai tersebut mengerjakan pekerjaannya sesuai dengan tugas dan tanggung jawab yang telah dibebankan. Hasil evaluasi tersebut dapat dijadikan sebagai *feedback* atau umpan balik kepada pegawai mengenai bagaimana kinerja mereka selama mereka bekerja.

Dengan adanya evaluasi penilaian kinerja pegawai, pihak pegawai maupun pihak instansi mendapatkan manfaat yang sebanding. Bagi pihak instansi, manfaat yang dirasakan adalah dapat menerima informasi yang jelas mengenai hasil pekerjaan yang telah dilakukan oleh pegawai. Selain itu juga pihak instansi dapat mendata pegawai mana saja yang bekerja sesuai dengan standar kinerja dan juga pegawai yang bekerja dibawah standar kinerja. Sedangkan bagi pihak pegawai, manfaat yang dirasakan adalah dapat mengetahui apa kelebihan serta kekurangan yang dimiliki oleh pegawai tersebut sehingga pegawai tersebut dapat memperbaiki kelemahannya dan dapat meningkatkan kelebihannya. Selain itu juga pihak pegawai dapat meningkatkan motivasi kerja agar dapat bekerja lebih baik lagi untuk kedepannya.

Melihat seberapa besar manfaat yang dirasa dengan adanya evaluasi penilaian kinerja karyawan, berbagai instansi pemerintah maupun swasta mulai mengadakan evaluasi penilaian kinerja pegawai. Salah satu instansi pemerintah yang mengadakan evaluasi penilaian kinerja pegawai adalah KPP PMA Dua. Kantor Pelayanan Pajak Penanaman Modal Asing Dua (KPP PMA Dua) merupakan salah satu kantor pajak yang beroperasi dalam lingkup Kantor Wilayah DJP Jakarta Khusus yang bertugas sebagai tempat pendaftaran dan pelaporan usaha oleh pemilik usaha yang bergerak pada bidang Industri Logam dan Mesin. KPP PMA Dua memiliki 10 divisi yang bertugas, Divisi Subbagian Umum dan Kepatuhan Internal, Divisi Pelayanan, Divisi Pengolahan Data dan Informasi (PDI), Divisi Pengawasan dan Konsultasi I (Waskon I), Divisi Pengawasan dan Konsultasi II (Waskon II), Divisi Pengawasan dan Konsultasi III (Waskon III), Divisi Pengawasan dan Konsultasi IV (Waskon IV), Divisi Penagihan, Divisi Pemeriksaan, serta Kelompok Jabatan Fungsional.

Kelompok Jabatan Fungsional sendiri terbagi menjadi dua kelompok, yaitu Fungsional Pemeriksa Pajak yang berisi 64 anggota serta Fungsional Penilai yang berisi 2 anggota. Tugas pokok dari Fungsional Pemeriksa Pajak yaitu untuk menguji kepatuhan Wajib Pajak (WP) dengan memeriksa bukti transaksi yang dilaporkan oleh Wajib Pajak serta melaksanakan pemeriksaan. Sedangkan tugas pokok dari fungsional Penilai yaitu menilai harga aset yang dimiliki oleh Wajib Pajak itu sendiri.

Dalam pelaksanaannya, Fungsional Pemeriksa Pajak terbagi atas 8 kelompok dengan masing-masing kelompok memiliki 1 ketua kelompok dan 1 kelompok berisi 3-4 tim. Masing-masing tim sendiri berisi 1 ketua tim dan 1 anggota tim. Dengan banyaknya pegawai Fungsional Pemeriksa Pajak, tentu butuh evaluasi dan penilai masing-masing pegawai agar kinerja dari tiap pegawai dapat optimal sesuai dengan kapabilitas. Maka dari itu KPP PMA Dua membentuk sistem penilaian anggota Fungsional Pemeriksa Pajak berdasarkan kinerja masing-masing pegawai Fungsional Pemeriksa Pajak. Akan tetapi dengan adanya inovasi baru tentu muncul permasalahan baru yang dihadapi, yaitu KPP PMA Dua kesulitan dalam melakukan penilaian pegawai Fungsional Pemeriksa Pajak. Permasalahan ini timbul karena banyaknya *variable* data yang digunakan dalam menentukan nilai masing-masing pegawai serta masih banyak data yang tersebar. Akibatnya, membutuhkan waktu yang lama serta ketelitian dalam mengolah data agar hasil dari pengolahan data tersebut tepat.

Data mining adalah sebuah metode yang digunakan untuk menggali dan mengumpulkan informasi penting dari kumpulan data yang lebih besar yang dapat digunakan dalam mengambil suatu keputusan. *Data mining* memiliki beberapa metode yang digunakan, yaitu deskripsi, estimasi, prediksi, klasifikasi, *clustering* serta asosiasi. Pada penelitian ini, peneliti menerapkan metode klasifikasi dengan menggunakan algoritma *naïve bayes*. Klasifikasi dapat diartikan sebagai salah satu bentuk pembelajaran data untuk mengekstrak model yang digunakan untuk memprediksi label kelas. *Naïve Bayes* sendiri merupakan salah satu metode klasifikasi yang juga menggunakan perhitungan statistik dengan menghitung kemungkinan kasus baru berdasarkan kemiripan dari kasus lama.

Penelitian mengenai evaluasi kinerja karyawan dengan menggunakan metode *naïve bayes* sebelumnya sudah dilakukan oleh (Wibowo & Hartati, 2016). Pada penelitian ini menggunakan data satpam PT. Garuda Merah Indonesia. Data yang didapat tersebut digunakan untuk membangun sistem untuk menilai kinerja satpam. Penelitian lainnya dilakukan oleh (Sari, et al., 2020). Penelitian ini menggunakan data karyawan PT. Pelita Wira Sejahtera untuk menilai kinerja karyawan tanpa adanya penilaian secara subjektif dan keterbatasan dalam mengontrol karyawan yang bekerja.

Agar data yang telah melalui proses *data mining* dapat dipahami oleh pihak KPP PMA Dua, diperlukan pembuatan *dashboard* untuk mempresentasikan hasil pengolahan data. Dengan adanya *dashboard*, maka informasi yang diberikan dapat dikomunikasikan dengan jelas dan efektif. *Dashboard* akan dibentuk untuk membantu organisasi dalam membentuk kebijakan serta membantu dalam pengambilan keputusan strategis.

Berdasarkan latar belakang di atas, peneliti bermaksud melakukan penelitian dengan judul “Analisis Hasil Evaluasi Kinerja Kelompok Jabatan Fungsional Pemeriksa Pajak Kantor Pelayanan PMA Dua Menggunakan Metode *Naïve Bayes*”. Hasil analisis yang telah didapatkan kemudian divisualisasikan dalam bentuk *dasboard* yang diharapkan dapat memudahkan jajaran KPP PMA Dua dalam

melihat kinerja Fungsional Pemeriksa Pajak yang sudah dikelompokkan berdasarkan nilai yang diraih.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, didapatkan rumusan masalah sebagai berikut:

1. Bagaimana hasil penerapan algoritma *Naïve Bayes* dalam mengukur kinerja Kelompok Jabatan Fungsional Pemeriksa Pajak?
2. Bagaimana hasil pengukuran *dashboard* dengan menggunakan metode *System Usability Scale* dalam membantu *top management* mengambil keputusan?

1.3 Tujuan

Tujuan dilakukan analisis hasil evaluasi kinerja Kelompok Jabatan Fungsional Pemeriksa KPP PMA Dua menggunakan Metode *Naïve Bayes*, di antaranya:

1. Menganalisis hasil evaluasi kinerja Kelompok Jabatan Fungsional menggunakan metode *Naïve Bayes*.
2. Menganalisis hasil pengukuran *dashboard* berdasarkan metode *System Usability Scale* dalam bentuk visualisasi data.

1.4 Manfaat

Dengan adanya penelitian ini, penulis mendapat beberapa manfaat, yaitu :

1. Penulis dapat memberikan masukan yang dapat digunakan oleh KPP PMA Dua seputar evaluasi kinerja Kelompok Jabatan Fungsional.
2. Mengimplementasikan metode *Naïve Bayes* yang dapat digunakan oleh KPP PMA Dua seputar evaluasi kinerja Kelompok Jabatan Fungsional.

1.5 Batasan Masalah

Adapun batasan-batasan masalah yang terdapat dalam penelitian ini yaitu:

1. Ruang lingkup dari penelitian ini dibatasi pada divisi Kelompok Jabatan Fungsional Pemeriksa Pajak.
2. Data yang digunakan pada penelitian ini adalah data kinerja Kelompok Jabatan Fungsional Pemeriksa Pajak tahun 2019 dan tahun 2020.
3. Penelitian ini tidak bertujuan untuk memberikan afirmasi atau penegasan bahwa pegawai yang mendapatkan nilai tertentu dari sistem penilaian merepresentasikan kinerja sesungguhnya dari pegawai tersebut.
4. Pengolahan data menggunakan Python dengan *tools* Google Colaboratory, dan data disimpan menggunakan Google Sheet.
5. Pembuatan *Dashboard* dilakukan menggunakan Google Data Studio.

1.6 Sistematika Pembahasan

Sistematika penulisan laporan menunjukkan gambaran dan struktur penulisan dari penelitian ini.

BAB 1 PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang penelitian, rumusan masalah, tujuan dari pelaksanaan penelitian, manfaat dari penelitian, batasan masalah dan sistematika pembahasan dari penelitian

BAB 2 LANDASAN KEPUSTAKAAN

Bab ini menjelaskan mengenai landasan teori serta referensi literatur yang digunakan sebagai dasar dan pendukung proses penelitian. Pada bab ini juga akan menjelaskan mengenai *tools* yang akan digunakan untuk mendukung penelitian ini serta pembahasan penelitian-penelitian sebelumnya yang dijadikan referensi dalam penelitian ini.

BAB 3 METODOLOGI

Bab ini menjelaskan mengenai langka-langkah yang dilakukan oleh penulis mulai dari awal hingga akhir penelitian. Selain itu juga bab ini memuat objek penelitian, metode atau teknik yang akan digunakan dalam proses penelitian.

BAB 4 IMPLEMENTASI

Bab ini memuat uraian mengenai proses penelitian yang dilakukan oleh penulis, yaitu tahap *pre-processing* serta penerapan metode *naïve bayes*.

BAB 5 HASIL DAN PEMBAHASAN

Pada bab ini akan menjelaskan mengenai hasil evaluasi kinerja berdasarkan metode *naïve bayes* serta pembahasan dari hasil yang telah didapatkan.

BAB 6 PENUTUP

Bab ini berisi mengenai kesimpulan serta saran yang telah dihasilkan dari penelitian ini. Diharapkan dari penelitian ini dapat berguna untuk penelitian selanjutnya.

BAB 2 LANDASAN KEPUSTAKAAN

2.1 Kajian Pustaka

Penelitian mengenai analisis kinerja menggunakan metode *Naïve Bayes* sebelumnya dilakukan oleh Viny Novika Sari pada tahun 2020. Penelitian dengan berjudul Analisis dan Penerapan Algoritma *Naïve Bayes* untuk Evaluasi Kinerja Karyawan pada PT. Pelita Wira Sejahtera ini bertujuan untuk menilai kinerja karyawan tanpa adanya subjektivitas dalam penilaian kinerja karyawan. Penelitian ini melibatkan 154 data dan setelah melakukan *cleaning data* data mengalami perubahan menjadi 149 data. Hasil dari penelitian ini ditemukan bahwa persentasi hasil akurasi terbesar diperoleh dengan menggunakan *Training Set* dengan persentasi akurasi sebesar 95.302%.

Penelitian selanjutnya yang dapat dijadikan sebagai rujukan dilakukan oleh Adityo Permana Wibowo pada tahun 2016. Penelitian dengan berjudul Sistem Klasifikasi Kinerja Satpam menggunakan Metode *Naïve Bayes Classifier* ini bertujuan untuk membentuk sistem untuk menilai kinerja satpam. Penelitian ini melibatkan 39 data uji dan data diambil dari tes kinerja yang terdiri dari kemampuan, kepribadian dan ketrampilan. Hasil dari penelitian ini menghasilkan presentasi kebenaran sebesar 92.31% dengan presentases kinerja baik sebesar 20,51%.

Penelitian berikutnya mengenai evaluasi penilaian kinerja pegawai dilakukan oleh Ikhsan Romli dan Boy Manggala Putra pada tahun 2021. Penelitian dengan berjudul Evaluasi Penilaian Kinerja Dalam Klasifikasi *Data Mining* Dengan Metode *Naïve Bayes* ini bertujuan untuk menilai kinerja satpam PT. Berkas Sinar Sentosa dan PT. Panin Bank Tbk. Penelitian ini melibatkan 63 dataset, 10 atribut dan 2 label serta menggunakan *tools* Rapidminer. Hasil dari penelitian menunjukkan presentasi akurasi model *naïve bayes* sebesar 85.71%.

2.2 Profil KPP PMA Dua

Kantor Pelayanan Pajak Penanaman Modal Asing Dua (KPP PMA Dua) merupakan salah satu kantor pajak yang beroperasi dalam lingkup Kantor Wilayah DJP Khusus yang berdiri sejak tahun 1997. KPP PMA Dua bertugas untuk mengoptimalkan tugas pengamanan penerimaan serta meningkatkan pelayanan kepada pemilik usaha yang bergerak pada bidang industri Logam dan Mesin.

Untuk mengoptimalkan pelayanan kepada Wajib Pajak, KPP PMA Dua memiliki 10 divisi yang bertugas, yaitu:

1. Subbagian Umum dan Kepatuhan Internal
Subbagian Umum dan Kepatuhan Internal bertugas melakukan urusan kepegawaian, keuangan, tata usaha serta rumah tangga.
2. Seksi Pelayanan

Seksi Pelayanan bertugas melakukan penetapan dan penerbitan produk hukum perpajakan, pengadministrasian dokumen dan berkas perpajakan, penerimaan dan pengolahan Surat Pemberitahuan, serta penerimaan surat lainnya, penyuluhan perpajakan, pelaksanaan registrasi Wajib Pajak serta melakukan kerjasama perpajakan.

3. Seksi Pengolahan Data dan Informasi (PDI)

Seksi Pengolahan Data dan Informasi bertugas melakukan pengumpulan, pencarian, dan pengolahan data, pengamatan potensi perpajakan, penyajian informasi perpajakan, perekaman dokumen perpajakan, pelayanan dukungan teknis komputer, pemantauan aplikasi *e-SPT* dan *e-Filing*, serta penyiapan laporan kinerja.

4. Seksi Pengawasan dan Konsultasi I (Waskon I)

5. Seksi Pengawasan dan Konsultasi II (Waskon II)

6. Seksi Pengawasan dan Konsultasi III (Waskon III)

7. Seksi Pengawasan dan Konsultasi IV (Waskon IV)

8. Divisi Penagihan

Seksi Penagihan bertugas melakukan urusan penatausahaan piutang pajak, penundaan dan angsuran tunggakan pajak, penagihan aktif, usulan penghapusan piutang pajak, serta penyimpanan dokumen-dokumen penagihan.

9. Divisi Pemeriksaan

Seksi Pemeriksaan bertugas melakukan penyusunan rencana pemeriksaan, pengawasan pelaksanaan aturan pemeriksaan, penerbitan dan penyaluran Surat Perintah Pemeriksaan Pajak serta administrasi pemeriksaan perpajakan lainnya.

10. Kelompok Jabatan Fungsional

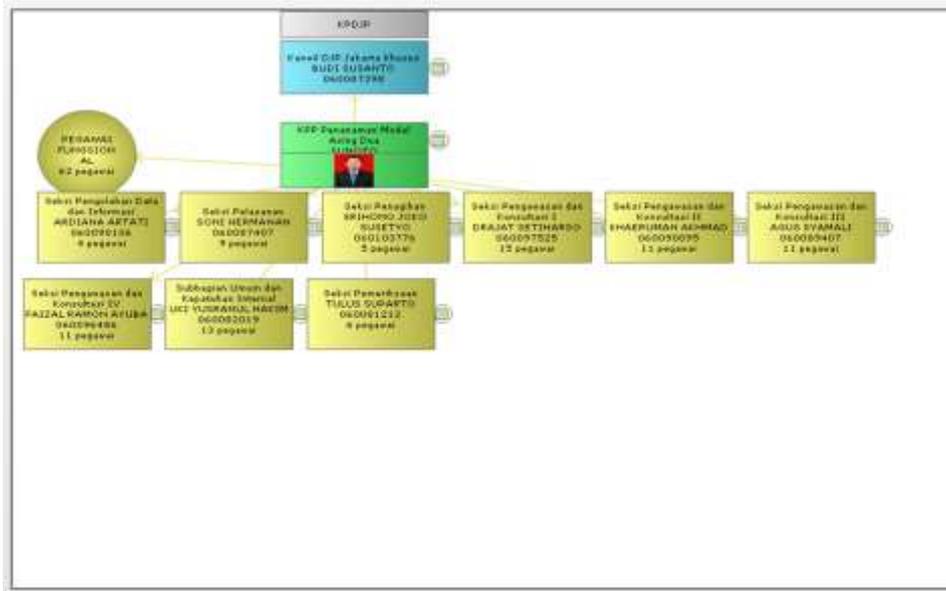
2.2.1 Visi

Menjadi institusi pemerintah yang menyelenggarakan Sistem Administrasi Perpajakan Modern yang efektif, efisien dan dipercaya masyarakat dengan integritas dan profesionalisme yang tinggi.

2.2.2 Misi

Menghimpun penerimaan pajak negara berdasarkan undang-undang perpajakan yang mampu mewujudkan kemandirian pembiayaan APBN melalui sistem administrasi perpajakan yang efisien dan efektif.

2.2.3 Struktur Organisasi



Gambar 2.1 Struktur Organisasi KPP PMA Dua

Sebagai sebuah organisasi, KPP PMA Dua tentu memiliki struktur organisasi yang berfungsi untuk memberikan tugas dan tanggung jawab setiap unsur organisasi sesuai yang terletak didalam susunan dalam struktur organisasi. KPP PMA Dua dibawah langsung oleh Kanwil DJP Jakarta Khusus dengan bapak Budi Susanto sebagai Kepala Kanwil. KPP PMA Dua dipimpin oleh bapak Sunoto sebagai Kepala Kantor. KPP PMA Dua membawahi 10 seksi dan Fungsional yang kedudukannya sejajar dengan Kepala Kantor.

2.3 Data Mining

Data mining merupakan suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Data mining juga diartikan sebagai proses *logic* yang berfungsi untuk menggali dari data yang ada untuk mendapatkan data yang dapat digunakan.

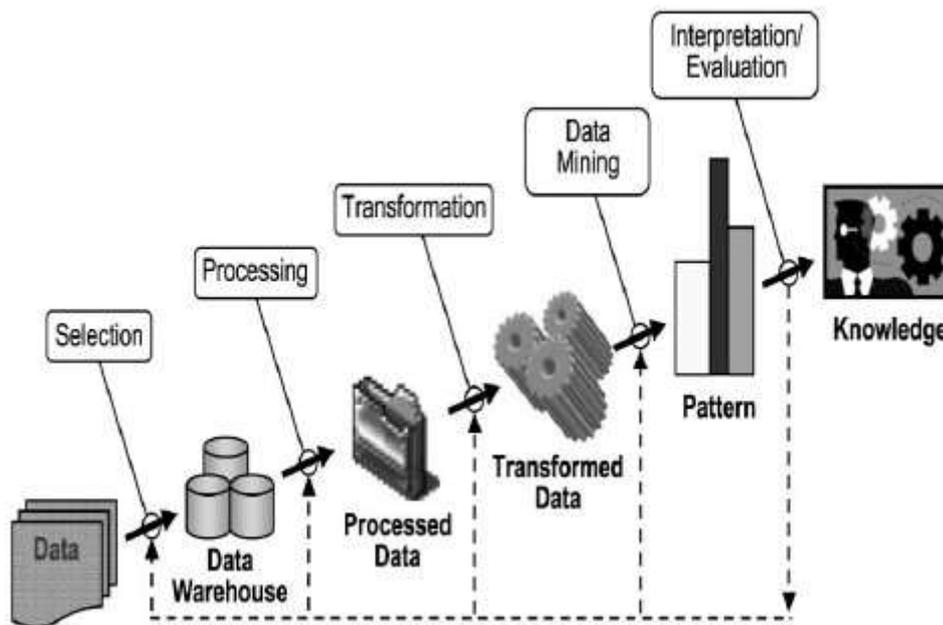
2.3.1 Tahapan Data Mining

Dalam menjalankan proses data mining, tentu ada beberapa tahapan yang perlu dijalankan agar mendapatkan hasil yang terbaik tanpa suatu kesalahan dalam mengolah data. Menurut Fayyad didalam jurnal yang berjudul Knowledge Discovery and Data Mining: Towards a Unifying Framework, ada 5 tahap yang perlu dilakukan, yaitu :

1. *Data selection*

Menyeleksi data terlebih dahulu dari kumpulan data sebelum memasuki tahap penggalian informasi. Data yang telah terseleksi yang akan digunakan dalam *data mining*, akan disimpan dalam berkas yang terpisah.

2. *Preprocessing*
Sebelum melaksanakan *data mining*, perlu melakukan tahapan *preprocessing*, yaitu sebuah tahap yang digunakan untuk membuang data yang terduplikasi, memperbaiki kesalahan dalam data serta memeriksa data yang masih inkonsisten.
3. *Transformation*
Transformation merupakan tahapan sebelum *data mining*, dimana data yang telah di *preprocessing* akan ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk *data mining*.
4. *Data mining*
Data mining merupakan proses yang digunakan untuk mencari informasi dalam data yang sudah melalui tahapan data selection hingga transformation. *Data mining* menggunakan metode atau algoritma yang tepat sesuai dengan tujuan dilakukannya proses *data mining*.
5. *Interpretation*
Interpretation merupakan proses pemeriksaan mengenai informasi yang ditemukan dalam *data mining* untuk iterasi lebih lanjut. Dalam tahap ini juga melibatkan visualisasi data. Tahapan ini juga biasa disebut *Evaluation*.



Gambar 2.2 Tahapan Data Mining

Sumber: Fayyad (1996)

2.3.2 Metode Data Mining

Dalam melakukan proses *data mining*, terdapat beberapa algoritma atau metode yang dapat digunakan, yaitu (Larose, 2005):

1. Deskripsi

Deskripsi merupakan metode *data mining* yang bekerja dengan memberikan gambaran atau deskripsi pada sekumpulan data untuk memahami lebih lanjut karakteristik data yang sedang diamati.

2. Estimasi

Estimasi merupakan metode yang bekerja dengan menerka atau menebak nilai yang belum diketahui. Contoh sederhana dari metode estimasi yaitu menerka gaji yang dimiliki oleh seorang karyawan yang belum diketahui informasinya.

3. Prediksi

Prediksi merupakan metode yang bekerja dengan menemukan pola suatu data dan menggunakan pola tersebut untuk memprediksi atau memperkirakan nilai yang akan datang. Contoh sederhana dari metode ini yaitu menebak prakiraan cuaca keesokan harinya.

4. Klasifikasi

Klasifikasi merupakan metode yang bekerja dengan membagi data menjadi beberapa kelas atau kelompok. Jika terdapat data baru yang tidak diketahui kelasnya, maka data tersebut dapat diperkirakan kelasnya berdasarkan kelas – kelas yang tersedia.

5. *Clustering*

Clustering merupakan metode yang bekerja dengan membagi data menjadi beberapa kelompok yang memiliki tingkat similaritas atau kesamaan yang tinggi berdasarkan karakteristik tertentu. Yang membedakan *clustering* dengan klasifikasi adalah pada *clustering* sendiri tidak menggunakan variable target.

6. Asosiasi

Asosiasi merupakan metode yang bekerja dengan menemukan dan mengidentifikasi kumpulan atribut yang muncul bersamaan berdasarkan kombinasi dari data. Contoh penggunaan metode berikut yaitu FP-Growth.

2.4 Klasifikasi

Klasifikasi merupakan metode data mining yang bekerja dengan memasukkan objek data kedalam kelas yang telah tersedia berdasarkan kelas yang tersedia untuk memprediksi label kelas. Metode klasifikasi terdiri dari dua proses, yaitu tahap pembelajaran serta tahap klasifikasi. Tahap pembelajaran bertujuan untuk membentuk model klasifikasi, sedangkan tahap klasifikasi bertujuan untuk memprediksi kelas dari sebuah data berdasarkan model klasifikasi dari tahap pembelajaran.

Klasifikasi sendiri dibagi menjadi beberapa algoritma, yaitu :

1. *Decision Tree*

Decision Tree merupakan metode klasifikasi yang bekerja dengan mengubah data atau fakta menjadi sebuah pohon keputusan yang berisi aturan. *Decision Tree* juga digunakan untuk membagi kumpulan data yang besar menjadi data yang lebih kecil dan memperlihatkan variabel tujuannya.

2. *Naïve Bayes*

Naïve Bayes merupakan metode klasifikasi yang biasa digunakan untuk memprediksi probabilitas suatu kelas. *Naïve Bayes* sendiri diambil dari teorema Bayes yang memiliki asumsi independensi yang kuat.

3. *K-Nearest Neighbor*

K-Nearest Neighbor merupakan metode klasifikasi berdasarkan data pembelajaran yang memiliki jarak paling dekat dengan objek tersebut. *K-Nearest Neighbor* mengklasifikasikan objek baru berdasarkan data dalam training data.

4. *Support Vector Machine*

Support Vector Machine merupakan metode klasifikasi yang bekerja dengan cara memetakan data ke dalam vektor untuk mencari hyperplane dengan memaksimalkan jarak antar kelas.

2.5 *Naïve Bayes*

Naïve Bayes adalah metode klasifikasi berbasis probabilitas yang digunakan untuk menghitung probabilitas atau kemungkinan yang akan datang dengan cara menjumlahkan frekuensi dan kombinasi nilai dari data yang diberikan. Chairul Fadhlán, et al (2018) mengungkapkan bahwa *Naïve Bayes* merupakan teknik prediksi berbasis probabilitas sederhana berdasarkan pada penerapan teorema Bayes dengan asumsi independensi yang kuat. *Naïve Bayes* termasuk ke dalam pembelajaran *supervised*, sehingga algoritma *Naïve Bayes* membutuhkan data awal yang digunakan tahap pembelajaran (*data training*) untuk mengambil keputusan.

Naïve Bayes merupakan teknik perhitungan teorema Bayes yang paling sederhana. Teorema ini menjelaskan jika terdapat dua kejadian yang saling terpisah, maka akan dirumuskan sebagai berikut :

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)} \quad (2.10)$$

Keterangan :

x = Data class yang belum diketahui

c = Hipotesis data

$P(c | x)$ = Probabilitas kemunculan kelas c dengan label kelas x

$P(x | c)$ = Probabilitas kemunculan kelas x dengan label kelas c

$P(x)$ = Probabilitas kemunculan kelas x

$P(c)$ = Probabilitas kemunculan kelas c

2.6 *K-Fold Cross Validation*

Cross Validation merupakan salah satu teknik validasi yang digunakan untuk mengevaluasi model *machine learning* yang bekerja dengan cara membagi data

menjadi dua bagian, bagian pertama digunakan untuk melatih model dan bagian kedua digunakan untuk menguji kinerja model tersebut (Bengio & Grandvalet, 2004). *Cross Validation* memiliki cara kerja yang mudah dipahami dan mudah diterapkan, sehingga banyak digunakan untuk mengevaluasi model *data mining*.

K-Fold Cross Validation merupakan salah satu bentuk dari *cross validation*. Metode *k-fold cross validation* bekerja dengan membagi data ke dalam beberapa bagian yang biasa disebut sebagai *fold*. Didalam *fold* tersebut memiliki jumlah data yang sama. Kemudian masing-masing *fold* akan dibagi menjadi data *training* dan data *test* selama berlangsungnya *k* iterasi, dimana satu *fold* akan dipilih menjadi data *test* dan sisa *k-1 fold* dijadikan sebagai data *training* (Refaeilzadeh, et al, 2009). Gambar 2.2 memuat bagaimana cara kerja dari metode *K-Fold Cross Validation*.

K-Fold Cross-Validation

Percobaan 1	Test	Train	Train	Train	Train
Percobaan 2	Train	Test	Train	Train	Train
Percobaan 3	Train	Train	Test	Train	Train
Percobaan 4	Train	Train	Train	Test	Train
Percobaan 5	Train	Train	Train	Train	Test

Gambar 2.3 Konsep K-Fold Cross Validation

2.7 Imbalanced Dataset

Imbalanced Dataset merupakan sebuah permasalahan didalam model klasifikasi dimana kelas pada dataset memiliki jumlah yang tidak seimbang, sehingga mengakibatkan hasil prediksi pada klasifikasi cenderung mengarah kepada kelas yang memiliki jumlah terbanyak (Untoro & Buliali, 2018). Untuk mengatasi permasalahan *imbalanced dataset*, dapat menerapkan metode *undersampling* serta *oversampling*. *Undersampling* bekerja dengan cara mengurangi data dari kelas yang memiliki jumlah terbanyak agar kelas didalam dataset memiliki jumlah data yang seimbang. Sedangkan *oversampling* bekerja dengan cara menambahkan data *sample* ke dalam kelas yang memiliki jumlah data yang sedikit.

2.8 Dashboard

Dashboard merupakan sebuah alat yang digunakan untuk mempresentasikan informasi penting dalam tampilan visual agar informasi yang disampaikan dapat

tersampaikan kepada penggunanya (Dewi, et al., 2015). Selain itu, *dashboard* juga dapat digunakan sebagai alat untuk memonitor kinerja organisasi serta pegawainya. Tujuan utama dengan adanya *dashboard* adalah membantu pengguna dalam pengambilan keputusan dengan tepat dari data yang telah disediakan. Agar pengguna dapat memahami informasi yang disampaikan dengan jelas, maka *dashboard* yang dibentuk perlu ditampilkan dengan visualisasi yang menarik. Saat ini banyak *tools* yang dapat dipakai untuk membangun sebuah *dashboard* yang memiliki tampilan yang menarik diantaranya Google Data Studio, GoodData, Tableau, dll.



Gambar 2.4 Tampilan Google Data Studio

2.9 System Usability Scale (SUS)

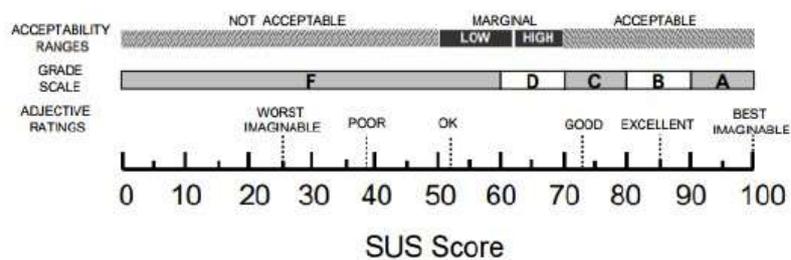
System Usability Scale merupakan sebuah alat ukur yang digunakan untuk mengukur tingkat *usability* atau tingkat kepuasan pengguna terhadap sistem yang telah dibangun. SUS dapat digunakan untuk mengevaluasi berbagai macam produk seperti sistem, website, aplikasi, dll. SUS terdiri dari 10 buah pertanyaan yang kemudian akan diberikan kepada responden untuk dinilai. Masing-masing pertanyaan terdapat lima poin jawaban yang dimulai dari sangat tidak setuju (1) hingga sangat setuju (5). Pertanyaan untuk kuesioner SUS dapat dilihat pada Tabel 2.1.

Tabel 2.1 Pertanyaan *System Usability Scale*

No.	Pertanyaan	Skala
-----	------------	-------

1.	Saya pikir saya akan sering menggunakan sistem ini	1 s/d 5
2.	Saya rasa sistem ini terlalu kompleks	1 s/d 5
3.	Saya kira sistem ini mudah untuk digunakan	1 s/d 5
4.	Saya pikir saya akan membutuhkan bantuan teknis untuk menggunakan sistem ini	1 s/d 5
5.	Saya kira ada berbagai fungsi dalam sistem ini terintegrasi dengan baik	1 s/d 5
6.	Saya pikir ada terlalu banyak ketidaksesuaian dalam sistem ini	1 s/d 5
7.	Saya bisa bayangkan bahwa kebanyakan orang akan mudah mempelajari dan menggunakan sistem ini	1 s/d 5
8.	Saya kira sistem ini sangat rumit untuk digunakan	1 s/d 5
9.	Saya merasa percaya diri dalam menggunakan sistem ini	1 s/d 5
10.	Saya perlu belajar banyak hal sebelum saya bisa memulai menggunakan sistem ini	1 s/d 5

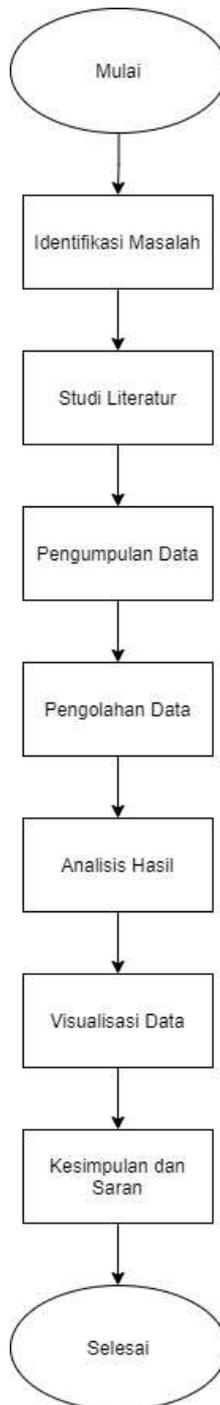
Sebelum menghitung jumlah skor yang diraih, terdapat beberapa aturan dalam menghitung jumlah skor. Pada pertanyaan dengan nomor ganjil (1, 3, 5, 7, 9), nilai yang telah dipilih oleh responden kemudian dikurangi dengan 1. Sedangkan pada pertanyaan dengan nomor genap (2, 4, 6, 8, 10), maka kurangkan 5 dengan nilai yang telah dipilih responden. Untuk memperoleh total nilai SUS, hasil jawaban responden dijumlahkan kemudian dikali dengan 2,5. Kemudian skor akhir SUS didapatkan dari rata-rata jawaban semua responden. Setelah mendapatkan skor akhir SUS, skor tersebut dibandingkan pada Gambar 2.4 untuk mengukur seberapa baik sistem yang telah dibangun berdasarkan jawaban dari responden.



Gambar 2.5 Skala Penilaian SUS

BAB 3 METODOLOGI

Dari latar belakang yang diangkat pada penelitian ini, didapatkan kerangka kerja yang menggambarkan kegiatan selama penelitian ini berlangsung. Kerangka ini juga dibuat berdasarkan metode yang dipilih. Berikut merupakan alur proses yang dijalankan dari awal hingga akhir penelitian:



Gambar 3.1 Diagram Tahapan Metodologi Penelitian

3.1 Identifikasi Masalah

Tahapan identifikasi masalah merupakan tahap awal yang dilakukan sebelum penelitian berlangsung. Tahap ini dijalankan untuk memahami permasalahan yang dialami oleh *stakeholder* dalam hal ini yaitu pihak KPP PMA Dua. Tahap ini penulis bekerja sama dengan *stakeholder* untuk identifikasi masalah yang dialami oleh KPP PMA Dua.

3.2 Studi Literatur

Tahapan selanjutnya yang dilaksanakan dalam penelitian ini adalah tahapan studi literatur. Tahap ini bertujuan untuk memperoleh pengetahuan dalam melaksanakan penelitian ini. Literatur yang digunakan pada penelitian kali ini yaitu buku, jurnal, penelitian sebelumnya serta *website* yang berhubungan dengan ilmu yang diperlukan. Studi literatur yang dipelajari untuk penelitian ini berkaitan dengan terkait *data mining*, algoritma *naïve bayes*, evaluasi *accuracy classifier*, *data visualization* dan *dashboard*, serta *usability testing* menggunakan *system usability scale*.

3.3 Pengumpulan Data

Pada tahap ketiga ini dilakukan pengumpulan data kepada pihak *stakeholder*. Pada tahap ini, penulis mendapatkan data primer melalui metode observasi serta wawancara kepada pihak *stakeholder*. Metode observasi dilakukan dengan mengamati langsung objek penelitian, yakni KPP PMA Dua. Sedangkan metode wawancara dilakukan dengan melakukan percakapan baik percakapan langsung maupun melalui media elektronik dengan narasumber berasal dari pihak KPP PMA Dua yaitu Kepala Seksi Pemeriksaan. Data yang diperoleh melalui metode observasi dan wawancara kepada pihak KPP PMA Dua merupakan data penilaian Kelompok Jabatan Fungsional Pemeriksa Pajak.

3.4 Pengolahan Data

Setelah tahapan pengumpulan data selesai dilakukan, tahapan yang selanjutnya dijalankan merupakan tahapan pengolahan data. Pada tahap ini, data melalui proses *data preprocessing* terlebih dahulu. Tahap *data preprocessing* dilakukan agar data yang baru diterima dari pihak *stakeholder* dapat disesuaikan dengan kebutuhan model *naïve bayes*. Dalam tahap *data preprocessing*, tahap yang dilakukan adalah *data selection* dan *data cleaning*. *Data selection* dilakukan untuk menentukan atribut serta jenis data yang akan digunakan. Sedangkan *data cleaning* dilakukan untuk menghilangkan nilai *null* atau *missing value* pada dataset. Setelah menyelesaikan tahap *data preprocessing*, kemudian data melalui proses penerapan model *naïve bayes*. Dalam tahap pengolahan data, *software* yang digunakan adalah Google Colaboratory dan Python untuk mempermudah tahap pengolahan data sesuai algoritma *naïve bayes*.

3.5 Pengujian dan Analisis Hasil

Selanjutnya penelitian dilanjutkan pada tahap pengujian dan analisis hasil. Model yang telah dibentuk akan diuji dengan menggunakan metode *K-Fold Cross Validation* serta *Accuracy Score*. Selain itu juga akan dilakukan tahap prediksi data untuk melihat seberapa baik model *naïve bayes* yang telah dibangun. Setelah data berhasil melalui tahap pengujian, maka dilanjutkan dengan analisis hasil berupa hasil nilai yang sudah disesuaikan berdasarkan algoritma *naïve bayes* untuk mendapatkan informasi yang akan digunakan oleh *stakeholder* untuk mengambil keputusan.

3.6 Visualisasi Hasil

Setelah melalui tahap pengujian dan analisis hasil, dilanjutkan dengan tahapan visualisasi hasil. Dalam tahapan ini, penulis menyampaikan informasi yang telah didapatkan dari model *naïve bayes* dengan membangun sebuah *dashboard*. Tahapan ini bertujuan untuk membantu pihak *stakeholder* dalam mengambil keputusan yang lebih baik kedepannya terkait dengan Fungsional Pemeriksa Pajak. Dalam tahapan ini juga dilakukan pengujian *dashboard* dengan menggunakan kuesioner *System Usability Scale (SUS)* untuk melihat bagaimana kepuasan *stakeholder* terhadap *dashboard* yang telah dibangun. Pembuatan *dashboard* dilakukan dengan memanfaatkan *software* Google Data Studio.

3.7 Kesimpulan dan Saran

Setelah menyelesaikan semua tahapan penelitian, tahapan terakhir yang dilakukan adalah tahapan kesimpulan dan saran. Di dalam tahap ini, penulis melakukan penarikan kesimpulan berdasarkan rumusan masalah dan hasil yang diperoleh pada penelitian kali ini. Penulis juga akan memberikan jawaban bagaimana kinerja dari Fungsional Pemeriksa Pajak. Selain itu penulis memberikan saran kepada pihak KPP PMA Dua berdasarkan kesimpulan penelitian dan juga masukan bagi pihak lain yang akan melakukan penelitian dengan topik yang sejenis.

BAB 4 IMPLEMENTASI

Pada bagian ini akan membahas mengenai implementasi yang telah dilakukan pada penelitian ini dengan menggunakan data yang telah didapatkan selama proses penelitian. Diawali dengan tahap pengumpulan data yang akan digunakan pada penelitian ini. Kemudian dilanjutkan kepada tahap *data preprocessing* yang digunakan untuk menghilangkan kolom yang tidak diperlukan. Selanjutnya, data yang sudah melalui fase *data preprocessing* akan digunakan pada proses *data mining* menggunakan *tools* Google Colab dan algoritma klasifikasi *naïve bayes*. Kemudian data yang sudah diproses kemudian dibuat visualisasi data dalam bentuk *dashboard* menggunakan *Google Data Studio*.

4.1 Pengumpulan Data

Tahap pengumpulan data dilakukan untuk mempelajari data yang dimiliki oleh objek penelitian, dalam hal ini berupa data penilaian kinerja fungsional pemeriksa pajak. Sebelum melakukan pengumpulan data, penulis terlebih dahulu mendapatkan izin dari pihak KPP PMA Dua. Data yang diminta adalah data yang sudah direkap oleh pihak KPP PMA Dua yang diambil dari sistem informasi KPP PMA Dua, dan sudah melalui tahap verifikasi data. Data yang diberikan oleh KPP PMA Dua berupa data pada tahun 2019 serta tahun 2020. Contoh data yang diberikan dapat dilihat pada gambar 4.1.

RANGKAI	BEL	NAMA PEMERIKSA	PENCAPAIAN SGP (50%)		SD (15%)		Kualitas Peng SP/ PPL (20%)		LIP (10%)		SOP (5%)		WIBUS SELESAI TERAT WAKTU (5%)		PRODUKSI DATA (5%)		TOTAL	RANGKAI BUKAH (NILAI)
			Capaian	Nilai	Capaian	Nilai	Capaian	Nilai	Capaian	Nilai	Capaian	Nilai	Capaian	Nilai	Capaian	Nilai		
1	11	IRHAM ALIA FACHRY	100.00%	0.94	8.87%	0.01	77.86%	0.12	100.00%	0.12	100.00%	0.06	100.00%	0.05	97.14%	0.01	0.92811	1
2	3	JUNI EDY GIAT SHAGA	100.00%	0.94	0.00%	0.00	100.00%	0.18	100.00%	0.11	100.00%	0.06	0.00%	0.00	96.15%	0.02	0.89031	2
3	8	ADLA HEBI	100.00%	0.94	8.87%	0.01	81.10%	0.11	87.14%	0.08	100.00%	0.06	0.00%	0.00	97.14%	0.01	0.88187	3
4	1	PARANTO	100.00%	0.94	0.00%	0.00	100.00%	0.18	94.29%	0.05	100.00%	0.08	0.00%	0.00	100.00%	0.00	0.83401	4
5	8	RUSLIWITO	100.00%	0.91	1.48%	0.00	10.80%	0.01	100.00%	0.12	100.00%	0.09	100.00%	0.05	98.87%	0.01	0.79971	5
6	2	RENAN WAHYU AI	100.00%	0.94	4.31%	0.01	87.41%	0.10	87.14%	0.04	48.58%	0.01	0.00%	0.00	100.00%	0.05	0.79885	6
7	8	SUTARNO ADRI WIDODO	100.00%	0.94	8.88%	0.01	7.98%	0.01	81.43%	0.08	22.88%	0.01	0.00%	0.00	98.88%	0.01	0.79619	7
8	1	ILHAM WICAKSONO	100.00%	0.94	1.38%	0.00	10.14%	0.01	97.14%	0.06	100.00%	0.06	0.00%	0.00	95.00%	0.01	0.68671	8
9	2	TUSUF CHOIR	100.00%	0.94	1.23%	0.00	0.00%	0.00	38.57%	0.04	100.00%	0.06	0.00%	0.00	96.47%	0.01	0.67419	9
10	11	TRI ARIWANINGIH	87.40%	0.93	1.80%	0.00	18.32%	0.01	100.00%	0.11	95.25%	0.05	100.00%	0.05	93.48%	0.01	0.66031	10
11	8	EDY SURIPTO	100.00%	0.94	0.00%	0.00	0.00%	0.00	94.29%	0.05	100.00%	0.06	0.00%	0.00	0.00%	0.00	0.65429	11
12	8	SULHAN TEDUH RAMBADI	100.00%	0.95	1.44%	0.00	7.76%	0.01	105.71%	0.11	100.00%	0.05	0.00%	0.00	95.80%	0.01	0.65359	12
13	8	ANDRIAMCO	100.00%	0.94	0.00%	0.00	0.00%	0.00	78.00%	0.08	4.71%	0.06	0.00%	0.00	98.81%	0.01	0.65086	13
14	2	YUNIM MATHO WICAKSONO	100.00%	0.94	1.88%	0.00	0.00%	0.00	98.57%	0.06	78.10%	0.04	0.00%	0.00	0.00%	0.01	0.64054	14
15	8	ILHAM WAHYUDIN	100.00%	0.93	1.02%	0.00	4.47%	0.01	92.86%	0.05	18.38%	0.01	50.00%	0.01	0.00%	0.00	0.60046	15
16	4	ANDI RISWANATO	83.01%	0.91	0.62%	0.00	45.94%	0.07	64.29%	0.06	100.00%	0.06	0.00%	0.00	14.59%	0.01	0.57462	16
17	1	CAHYO KUSUMA WARDANA	83.88%	0.98	0.82%	0.00	11.24%	0.01	100.00%	0.12	17.84%	0.01	35.20%	0.01	0.00%	0.00	0.54105	17
18	11	IRDA MUSUM	89.42%	0.91	17.00%	0.01	15.70%	0.01	84.29%	0.08	100.00%	0.06	0.00%	0.00	0.00%	0.00	0.53071	18
19	4	PERFY KIKUH WICAKSONO	88.00%	0.91	3.88%	0.01	90.91%	0.04	77.14%	0.08	88.84%	0.03	0.00%	0.00	0.00%	0.00	0.49740	19
20	8	ROHM DAMAZAN	85.80%	0.93	2.01%	0.00	0.00%	0.00	109.29%	0.11	88.58%	0.05	0.00%	0.00	98.81%	0.01	0.48079	20
21	2	IMELI SIKHANTO	84.00%	0.93	1.08%	0.00	14.17%	0.01	16.71%	0.01	67.40%	0.01	0.00%	0.00	0.00%	0.00	0.48071	21
22	8	PEH HINDONO	83.50%	0.94	0.00%	0.00	0.00%	0.00	84.29%	0.06	60.88%	0.01	40.00%	0.01	81.33%	0.01	0.44601	22
23	2	DHAWANG SRI WIDODO	86.00%	0.90	2.28%	0.00	0.00%	0.00	63.57%	0.00	66.99%	0.01	0.00%	0.00	17.14%	0.01	0.42702	23

Gambar 4.1 Data Fungsional Pemeriksa Pajak Tahun 2019

Data yang penulis terima dari *stakeholder* merupakan dataset rekap capaian dan nilai yang telah diraih oleh fungsional pemeriksa pajak selama satu tahun. Dataset ini memiliki 19 atribut serta 1 tambahan atribut yaitu Nilai Huruf. Atribut Nilai Huruf merupakan atribut tambahan yang dibuat berdasarkan hasil wawancara penulis dengan *stakeholder*, karena sebelumnya KPP PMA Dua sendiri belum pernah menggunakan atribut nilai Huruf dan lebih berfokus kepada atribut total nilai. Penilaian atribut kelas sendiri diambil dari konversi total nilai yang

didapatkan oleh fungsional pemeriksa pajak dan dikonversikan ke nilai huruf. Dari dataset tersebut, tidak semua atribut digunakan dan hanya menggunakan atribut capaian serta atribut nilai huruf. Atribut yang didapatkan dapat dilihat pada tabel 4.1, sedangkan atribut yang akan digunakan dapat dilihat pada tabel 4.2.

Tabel 4.1 Keseluruhan Atribut Dataset

No	Nama Atribut	Keterangan
1	Rank	Ranking yang diraih oleh fungsional pemeriksa pajak pada tahun ini
2	Kelompok	Kelompok yang sudah ditempati oleh fungsional pemeriksa pajak
3	Nama Pemeriksa	Nama lengkap fungsional pemeriksa pajak
4	Capaian Pencairan SKP	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian pencairan SKP
5	Nilai Pencairan SKP	Nilai yang didapatkan dari capaian pencairan SKP
6	Capaian RD	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian RD (<i>Refund Discrepancy</i>)
7	Nilai RD	Nilai yang didapatkan dari capaian RD
8	Capaian Koreksi Rugi SPT	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian Koreksi Rugi SPT
9	Nilai Koreksi Rugi SPT	Nilai yang didapatkan dari capaian koreksi rugi SPT
10	Capaian LHP	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian LHP
11	Nilai LHP	Nilai yang didapatkan dari capaian LHP
12	Capaian SKPKB Terbit	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian SKPKB Terbit
13	Nilai SKPKB Terbit	Nilai yang didapatkan dari capaian SKPKB Terbit
14	Capaian Riksus Selesai	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian Riksus Selesai
15	Nilai Riksus Selesai	Nilai yang didapatkan dari capaian riksus selesai
16	Capaian Produksi Data	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian Produksi Data

17	Nilai Produksi Data	Nilai yang didapatkan dari capaian produksi data
18	Total Nilai	Nilai akumulatif yang didapatkan oleh fungsional pemeriksa pajak
19	Rank Bulan Lalu	Rank yang diraih oleh fungsional pemeriksa pajak pada bulan sebelumnya

Tabel 4.2 Atribut Dataset yang digunakan

No	Nama Atribut	Keterangan
1	Nama Pemeriksa	Nama lengkap fungsional pemeriksa pajak
2	Capaian Pencairan SKP	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian pencairan SKP
3	Capaian RD	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian RD (<i>Refund Discrepancy</i>)
4	Capaian Koreksi Rugi SPT	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian Koreksi Rugi SPT
5	Capaian LHP	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian LHP
6	Capaian SKPKB Terbit	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian SKPKB Terbit
7	Capaian Riksus Selesai	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian Riksus Selesai
8	Capaian Produksi Data	Capaian yang diraih oleh fungsional pemeriksa pajak dalam penyelesaian Produksi Data
9	Nilai Huruf	Nilai akhir yang diterima oleh fungsional pemeriksa pajak berdasarkan total nilai yang didapat dari 7 nilai sebelumnya.

4.2 Data Preprocessing

Data Preprocessing merupakan sebuah cara yang dilakukan untuk mengubah data mentah menjadi informasi yang bermanfaat. Pada penelitian ini perlu melalui proses *data preprocessing* karena dataset tersebut memiliki atribut yang sangat banyak sehingga perlu dikurangi sesuai yang dibutuhkan untuk penerapan algoritma Naïve Bayes. Dalam penerapan *data preprocessing*, penulis menerapkan *import data*, *data cleaning* serta *data selection*.

4.2.2 Data Selection & Data Cleaning

Setelah melakukan proses *import data*, dilanjutkan dengan tahap *data selection*. *Data selection* merupakan proses untuk menentukan jenis data serta atribut yang akan digunakan pada proses *data mining*. Atribut data yang digunakan pada model *Naïve Bayes* adalah Nama Pemeriksa, Capaian Pencairan SKP, Capaian RD, Capaian Koreksi Rugi SPT, Capaian LHP, Capaian LHP, Capaian Riksus Selesai, Capaian Produksi Data dan Nilai Huruf. Sedangkan atribut lainnya yang tidak digunakan akan di *drop*. Pada tahap ini juga dilakukan tahap *data cleaning*, yaitu proses untuk mendeteksi dan menghilangkan data yang bernilai *null*. Untuk mengetahui berapa banyak data yang *null* pada masing-masing dataset digunakan fungsi *isnull().sum()*. Implementasi proses *data selection* dan *data cleaning* dapat dilihat pada Kode Program 4.2.

```
1 # Data 2019
2 # Rename Pencairan_SKP_Capaian
3 df_pemeriksa2019 = df_pemeriksa2019.rename
  (columns={"Pencairan_SKP_Capaian":
  "Pencairan_SKP_Capaian_(%)"})
4 # Rename RD_Capaian
5 df_pemeriksa2019 = df_pemeriksa2019.rename
  (columns={"RD_Capaian": "RD_Capaian_(%)"})
6 # Rename Koreksi_Rugi_SPT_Capaian
7 df_pemeriksa2019 = df_pemeriksa2019.rename
  (columns={"Koreksi_Rugi_SPT_Capaian":
  "Koreksi_Rugi_SPT_Capaian_(%)"})
8 # Rename LHP_Capaian
9 df_pemeriksa2019 = df_pemeriksa2019.rename
  (columns={"LHP_Capaian": "LHP_Capaian_(%)"})
10 # Rename SKPKB_Terbit_Capaian
11 df_pemeriksa2019 = df_pemeriksa2019.rename
  (columns={"SKPKB_Terbit_Capaian":
  "SKPKB_Terbit_Capaian_(%)"})
12 # Rename Rikus_Selesai_Capaian
13 df_pemeriksa2019 = df_pemeriksa2019.rename
  (columns={"Rikus_Selesai_Capaian":
  "Rikus_Selesai_Capaian_(%)"})
14 # Rename Produksi_Data_Capaian
15 df_pemeriksa2019 = df_pemeriksa2019.rename
  (columns={"Produksi_Data_Capaian":
  "Produksi_Data_Capaian_(%)"})
16 # Drop kolom yang tidak dibutuhkan
17 df_pemeriksa2019 = df_pemeriksa2019.drop("Rank", 1)
18 df_pemeriksa2019 = df_pemeriksa2019.drop("Kelompok", 1)
19 df_pemeriksa2019 = df_pemeriksa2019.drop("Pencairan_SKP_Nila
  i", 1)
20 df_pemeriksa2019 = df_pemeriksa2019.drop("RD_Nilai", 1)
21 df_pemeriksa2019 = df_pemeriksa2019.drop("Koreksi_Rugi_SPT_N
  ilai", 1)
```

```

22 df_pemeriksa2019 = df_pemeriksa2019.drop("LHP_Nilai", 1)
23 df_pemeriksa2019 = df_pemeriksa2019.drop("SKPKB_Terbit_Nilai", 1)
24 df_pemeriksa2019 = df_pemeriksa2019.drop("Rikus_Selesai_Nilai", 1)
25 df_pemeriksa2019 = df_pemeriksa2019.drop("Produksi_Data_Nilai", 1)
26 df_pemeriksa2019 = df_pemeriksa2019.drop("Rank_Bulan_Lalu", 1)
27 df_pemeriksa2019 = df_pemeriksa2019.drop("Total_Nilai", 1)
28 df_pemeriksa2019.info()

29 # Mengubah tipe data dari object menjadi float
30 df_pemeriksa2019["Pencairan_SKP_Capaian_(")] = df_pemeriksa2019["Pencairan_SKP_Capaian_(")].astype(str).astype(float)
31 df_pemeriksa2019["RD_Capaian_(")] = df_pemeriksa2019["RD_Capaian_(")].astype(str).astype(float)
32 df_pemeriksa2019["Koreksi_Rugi_SPT_Capaian_(")] = df_pemeriksa2019["Koreksi_Rugi_SPT_Capaian_(")].astype(str).astype(float)
33 df_pemeriksa2019["LHP_Capaian_(")] = df_pemeriksa2019["LHP_Capaian_(")].astype(str).astype(float)
34 df_pemeriksa2019["SKPKB_Terbit_Capaian_(")] = df_pemeriksa2019["SKPKB_Terbit_Capaian_(")].astype(str).astype(float)
35 df_pemeriksa2019["Rikus_Selesai_Capaian_(")] = df_pemeriksa2019["Rikus_Selesai_Capaian_(")].astype(str).astype(float)
df_pemeriksa2019["Produksi_Data_Capaian_(")] = df_pemeriksa2019["Produksi_Data_Capaian_(")].astype(str).astype(float)
36 df_pemeriksa2019.info()

37 # Mengecek missing value
38 df_pemeriksa2019.isnull().sum()

39 # Mengecek isi kolom terbaru setelah melakukan pre processing
40 df_pemeriksa2019

```

Kode Program 4.2 Data Selection & Data Cleaning

Struktur data sebelum melakukan *data selection* dan *data cleaning* dapat dilihat pada Gambar 4.3 dimana semua data masih berbentuk *object*. Maka dari itu, perlu dilakukan pengubahan tipe data dari masih berbentuk *object* menjadi *float*. Selain itu juga pada tahap ini dilakukan proses *drop* atribut yang tidak digunakan seperti atribut Rank, Kelompok, Pencairan_SKP_Nilai, RD_Ni;ai, Koreksi_Rugi_SPT_Nilai, LHP_Nilai, SKPKB_Terbit_Nilai, Rikus_Selesai_Nilai, Produksi_Data_Nilai, Total_Nilai dan Rank_Bulan_Lalu. Kemudian nama atribut yang akan digunakan diubah dengan fungsi *rename* untuk menambahkan persen (%) diakhir nama atribut tersebut. Hasil struktur data tahap *data selection* dapat dilihat pada Gambar 4.4.

```

Data columns (total 20 columns):
#   Column                               Non-Null Count  Dtype
---  -
0   Rank                                   64 non-null     object
1   Kelompok                               64 non-null     object
2   Nama_Pemeriksa                         64 non-null     object
3   Pencairan_SKP_Capaian                 64 non-null     object
4   Pencairan_SKP_Nilai                   64 non-null     object
5   RD_Capaian                             64 non-null     object
6   RD_Nilai                               64 non-null     object
7   Koreksi_Rugi_SPT_Capaian              64 non-null     object
8   Koreksi_Rugi_SPT_Nilai                64 non-null     object
9   LHP_Capaian                           64 non-null     object
10  LHP_Nilai                              64 non-null     object
11  SKPKB_Terbit_Capaian                  64 non-null     object
12  SKPKB_Terbit_Nilai                    64 non-null     object
13  Rikus_Selesai_Capaian                 64 non-null     object
14  Rikus_Selesai_Nilai                   64 non-null     object
15  Produksi_Data_Capaian                 64 non-null     object
16  Produksi_Data_Nilai                   64 non-null     object
17  Total_Nilai                           64 non-null     object
18  Rank_Bulan_Talu                       64 non-null     object
19  Nilai_Huruf                           64 non-null     object
dtypes: object(20)
memory usage: 10.1+ KB

```

Gambar 4.3 Struktur Dataset Penilaian Sebelum Data Selection

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 64 entries, 0 to 63
Data columns (total 9 columns):
#   Column                               Non-Null Count  Dtype
---  -
0   Nama_Pemeriksa                       64 non-null     object
1   Pencairan_SKP_Capaian_(%)            64 non-null     float64
2   RD_Capaian_(%)                       64 non-null     float64
3   Koreksi_Rugi_SPT_Capaian_(%)         64 non-null     float64
4   LHP_Capaian_(%)                     64 non-null     float64
5   SKPKB_Terbit_Capaian_(%)            64 non-null     float64
6   Rikus_Selesai_Capaian_(%)           64 non-null     float64
7   Produksi_Data_Capaian_(%)            64 non-null     float64
8   Nilai_Huruf                          64 non-null     object
dtypes: float64(7), object(2)
memory usage: 4.6+ KB

```

Gambar 4.4 Struktur Dataset Penilaian Setelah Data Selection

Setelah melakukan tahap *data selection*, kemudian dilanjutkan dengan tahap *data cleaning*. Untuk melihat apakah ada data yang *null* atau tidak pada dataset, digunakan fungsi *isnull().sum()*. Namun pada dataset tahun 2019 dan tahun 2020 tidak ada data *null* seperti yang ditunjukkan pada Gambar 4.5, sehingga tidak perlu menggunakan fungsi *dropna()*. Hasil data yang telah melalui proses *data selection* dan *data cleaning* dapat dilihat pada tabel 4.3.

```

Nama_Pemeriksa           0
Pencairan_SKP_Capaian_(%) 0
RD_Capaian_(%)          0
Koreksi_Rugi_SPT_Capaian_(%) 0
LHP_Capaian_(%)         0
SKPKB_Terbit_Capaian_(%) 0
Rikus_Selesai_Capaian_(%) 0
Produksi_Data_Capaian_(%) 0
Nilai_Huruf             0
dtypes: int64

```

Gambar 4.5 Struktur Dataset Penilaian Hasil Data Selection

Tabel 4.3 Data Hasil *Preprocessing*

Nama_ Pemeriksa	Pencairan_ SKP_ Capaian	RD_ Capaian	Koreksi_ Rugi_ SPT_ Capaian	LHP_ Capaian	SKPK B_ Terbit _ Capaian	Riksus _ Selesai_ Capaian	Produk si_ Data_ Capaian	Nilai_ Huruf
Irkham Aulia Fathony	120.0	8.45	120.0	120.0	120.0	100.0	50.0	A
Rudiyanto	114.48	1.36	120.0	120.0	120.0	100.0	66.67	A
Juni Eddy Giat Sinaga	120.0	0.05	120.0	112.86	120.0	0.0	42.86	A
...
Karmuji	0.49	1.44	0.00	75.71	0.49	0.0	7.69	E
Budi Rachmanto	3.54	0.03	1.44	50.0	8.54	0.0	0.0	E
Tri Budi Rahardjo	5.01	0.00	0.00	26.43	6.10	0.0	0.0	E

4.3 Penerapan *Naïve Bayes*

Tahapan Klasifikasi dilakukan setelah tahap *data preprocessing* selesai dilakukan. Data yang sudah sesuai dengan kebutuhan selanjutnya diolah menggunakan model klasifikasi dengan algoritma *Naïve Bayes* pada Google Colaboratory. Adapun hasil *output* yang diharapkan pada tahapan ini adalah hasil prediksi nilai huruf pada data tahun 2020 yang didapatkan dari *data training* pada data tahun 2019. Pada penelitian ini menerapkan *libraries* SMOTE karena dataset karyawan tahun 2019 merupakan *data imbalanced*, sehingga membutuhkan teknik *oversampling* untuk mengimbangi *data imbalanced* tersebut. Pada proses pelatihan model *naïve bayes*, dataset perlu dibagi menjadi data training serta data test dengan memanfaatkan fungsi *train_test_split* menggunakan library dari *sklearn.model_selection* dengan rasio pembagian data sebesar 80 % untuk data

training dan 20% untuk data *testing*. Implementasi model *naïve bayes* ditunjukkan pada Kode Program 4.3.

```
1 # tahun 2019
2 from sklearn.model_selection import train_test_split
3 from sklearn.naive_bayes import GaussianNB
4 from numpy import mean
5 from numpy import std
6
7 prediksiX = df_pemeriksa2019[['Pencairan_SKP_Capaian_(%)',
8 'RD_Capaian_(%)', 'Koreksi_Rugi_SPT_Capaian_(%)', 'LHP_Capai
9 an_(%)', 'SKPKB_Terbit_Capaian_(%)', 'Rikus_Selesai_Capaian_
10 (%)', 'Produksi_Data_Capaian_(%)']]
11 prediksiY = df_pemeriksa2019['Nilai_Huruf']
12
13 from imblearn.over_sampling import SMOTE
14 from sklearn.preprocessing import LabelEncoder
15
16 oversample = SMOTE()
17 prediksiX, prediksiY = oversample.fit_resample(prediksiX, pr
18 ediksiY)
19 X_train, X_test, y_train, y_test = train_test_split(prediksi
20 X, prediksiY, test_size = 0.2, random_state= 141)
21 gnb = GaussianNB()
22 nbtrain = gnb.fit(X_train, y_train)
23
24 y_pred = gnb.predict(X_test)
25 y_pred
26
27 np.array(y_test)
```

Kode Program 4.3 Penerapan Naïve Bayes

4.4 Hasil Akurasi Klasifikasi *Naïve Bayes*

Setelah menyelesaikan tahapan penyusunan model klasifikasi menggunakan *naïve bayes*, kemudian perlu dilakukan tahapan uji validasi untuk melihat bagaimana performa dari model klasifikasi *naïve bayes* yang dibentuk. Metode yang dipakai dalam proses uji validasi ini ada dua metode, yakni *Accuracy Score* dan *K-Fold Cross Validation*. Pada metode *accuracy score* menggunakan variable *y_pred* dan *y_test* sebagai variable pembanding dan library dari *sklearn metrics*. Sedangkan pada metode *K Fold Cross Validation* menggunakan library *KFold* serta penggunaan metode *cross_val_score* untuk menghitung nilai uji validasi. Implementasi pada Google Colab dapat dilihat pada Kode Program 4.4.

```

1 #Accuracy Score
2 from sklearn import metrics
3 print(metrics.accuracy_score(y_test, y_pred))

4 from sklearn.model_selection import KFold
5 kf1 = KFold(n_splits = 5, random_state=141, shuffle= True)
6 kf1

7 for train_index, test_index in kf1.split(prediksiX):
8     print(train_index, test_index)

9 #5Fold
10 from sklearn.model_selection import cross_val_score
11 scores = cross_val_score(gnb, prediksiX, prediksiY, cv=kf1,
12 scoring='accuracy')
13 print(scores.mean())

```

Kode Program 4.4 Hasil Akurasi *Naïve Bayes*

4.5 Prediksi Data

Tahap selanjutnya adalah melakukan prediksi data fungsional pemeriksa pajak tahun 2020 menggunakan model klasifikasi *naïve bayes* yang baru saja dibentuk. Sebelum melakukan tahap prediksi data, data tahun 2020 perlu di *import* ke Google Colab dan perlu melalui tahap *data preprocessing* untuk menyamakan atribut yang akan digunakan ada tahap model *naïve bayes* yang sudah dibentuk. Setelah melalui proses *data preprocessing*, kemudian dilanjutkan dengan pembentukan variable *nilaiX2020* untuk menyimpan atribut yang akan digunakan untuk memprediksi variable *nilai_huruf* pada data tahun 2020. Tahap prediksi data kemudian dilakukan dengan menggunakan fungsi *gnb.predict* untuk memprediksi nilai huruf data tahun 2020 menggunakan parameter variable *nilaiX2020*. Implementasi Prediksi Data dapat dilihat pada Kode Program 4.5. Hasil prediksi nilai huruf dapat dilihat pada Gambar 4.4.

```

1 #import data 2020
2 ws2020 = workbook.worksheet('Pemeriksa2020')

3 # get_all_values gives a list of rows
4 rows2020 = ws2020.get_all_values()

5 # Convert to a DataFrame and render.
6 df_pemeriksa2020 = pd.DataFrame.from_records(rows2020[1:], co
7 lumns=rows2020[0])

8 df_pemeriksa2020
9 #Struktur data
10 df_pemeriksa2020.info()

11 # Rename Pencairan_SKP_Capaian
12 df_pemeriksa2020 = df_pemeriksa2020.rename(columns={"Pencair
13 an_SKP_Capaian": "Pencairan_SKP_Capaian_("%)"})

14 # Rename RD_Capaian
15 df_pemeriksa2020 = df_pemeriksa2020.rename(columns={"RD_Capa
16 ian": "RD_Capaian_("%)"})

```

```

15 # Rename Koreksi_Rugi_SPT_Capaian
16 df_pemeriksa2020 = df_pemeriksa2020.rename(columns={"Koreksi
_Rugi_SPT_Capaian": "Koreksi_Rugi_SPT_Capaian_(%)"})

17 # Rename LHP_Capaian
18 df_pemeriksa2020 = df_pemeriksa2020.rename(columns={"LHP_Cap
aian": "LHP_Capaian_(%)"})

19 # Rename SKPKB_Terbit_Capaian
20 df_pemeriksa2020 = df_pemeriksa2020.rename(columns={"SKPKB_T
erbit_Capaian": "SKPKB_Terbit_Capaian_(%)"})

21 # Rename Rikus_Selesai_Capaian
22 df_pemeriksa2020 = df_pemeriksa2020.rename(columns={"Rikus_S
elesai_Capaian": "Rikus_Selesai_Capaian_(%)"})

23 # Rename Produksi_Data_Capaian
24 df_pemeriksa2020 = df_pemeriksa2020.rename(columns={"Produks
i_Data_Capaian": "Produksi_Data_Capaian_(%)"})

25 # Mengecek missing value
26 df_pemeriksa2020.isnull().sum()

27 # Drop kolom yang tidak dibutuhkan
28 df_pemeriksa2020 = df_pemeriksa2020.drop("Rank", 1)
29 df_pemeriksa2020 = df_pemeriksa2020.drop("Kelompok", 1)
30 df_pemeriksa2020 = df_pemeriksa2020.drop("Pencairan_SKP_Nila
i", 1)
31 df_pemeriksa2020 = df_pemeriksa2020.drop("RD_Nilai", 1)
32 df_pemeriksa2020 = df_pemeriksa2020.drop("Koreksi_Rugi_SPT_N
ilai", 1)
33 df_pemeriksa2020 = df_pemeriksa2020.drop("LHP_Nilai", 1)
34 df_pemeriksa2020 = df_pemeriksa2020.drop("SKPKB_Terbit_Nilai
", 1)
35 df_pemeriksa2020 = df_pemeriksa2020.drop("Rikus_Selesai_Nila
i", 1)
36 df_pemeriksa2020 = df_pemeriksa2020.drop("Produksi_Data_Nila
i", 1)
37 df_pemeriksa2020 = df_pemeriksa2020.drop("Rank_Bulan_Lalu",1
)
38 df_pemeriksa2020 = df_pemeriksa2020.drop("Total_Nilai", 1)
39 df_pemeriksa2020.info()

40 # Mengubah tipe data dari object menjadi float
41 df_pemeriksa2020["Pencairan_SKP_Capaian_(%)"] = df_pemeriksa
2020["Pencairan_SKP_Capaian_(%)"].astype(str).astype(float)
42 df_pemeriksa2020["RD_Capaian_(%)"] = df_pemeriksa2020["RD_Ca
paian_(%)"].astype(str).astype(float)
43 df_pemeriksa2020["Koreksi_Rugi_SPT_Capaian_(%)"] = df_pemeri
ksa2020["Koreksi_Rugi_SPT_Capaian_(%)"].astype(str).astype(f
loat)
44 df_pemeriksa2020["LHP_Capaian_(%)"] = df_pemeriksa2020["LHP_
Capaian_(%)"].astype(str).astype(float)
45 df_pemeriksa2020["SKPKB_Terbit_Capaian_(%)"] = df_pemeriksa2
020["SKPKB_Terbit_Capaian_(%)"].astype(str).astype(float)
46 df_pemeriksa2020["Rikus_Selesai_Capaian_(%)"] = df_pemeriksa
2020["Rikus_Selesai_Capaian_(%)"].astype(str).astype(float)
47 df_pemeriksa2020["Produksi_Data_Capaian_(%)"] = df_pemeriksa
2020["Produksi Data Capaian (%)"].astype(str).astype(float)

```

```

48 df_pemeriksa2020.info()
49 #tahap prediksi nilai_huruf
50 nilaiX2020 = df_pemeriksa2020[['Pencairan_SKP_Capaian_(%)',
    'RD_Capaian_(%)', 'Koreksi_Rugi_SPT_Capaian_(%)', 'LHP_Capai
    an_(%)', 'SKPKB_Terbit_Capaian_(%)', 'Rikus_Selesai_Capaian_
    (%)', 'Produksi_Data_Capaian_(%)']]
51 prediksi2020 = gnb.predict(nilaiX2020)
52 print(prediksi2020)

53 nama_2020 = df_pemeriksa2020['Nama_Pemeriksa']
54 nama_2020
55 nama_2020 = pd.DataFrame(nama_2020)
56 nilaiX2020.insert(0, 'Nama_Pemeriksa', nama_2020)
57 hasil_pred_akhir2020 = nilaiX2020
58 print(hasil_pred_akhir2020)
59 hasil_pred_akhir2020['Nilai_Huruf'] = prediksi2020
60 print(hasil_pred_akhir2020)

```

Kode Program 4.5 Prediksi Data Tahun 2020

4.6 Dashboard Kinerja Pegawai

Setelah melakukan tahap prediksi variable nilai_huruf, kemudian dilanjutkan tahap berikutnya yaitu penyusunan *dashboard* kinerja pegawai dengan menggunakan *tools* Google Data Studio. Data yang digunakan pada *dashboard* ini adalah data tahun 2019 dan data tahun 2020. Fungsi dari *dashboard* ini sendiri adalah untuk membandingkan bagaimana performa dari fungsional pemeriksa pajak pada tahun 2019 dan tahun 2020.

Sebelum menyusun *dashboard*, terlebih dulu melakukan tahap *export* data hasil *preprocessing* 2019 dan hasil prediksi 2020 dari Google Colab ke Google Sheet. Tahapan *export* dilakukan dengan menggunakan fungsi *set_with_dataframe* dari library Gspread. Adapun data yang di *export* merupakan dataframe tahun 2019 dan tahun 2020. Implementasi *export* data dapat dilihat pada Kode Program 4.6.

```

1 from gspread_dataframe import get_as_dataframe, set_with_data
  aframe

2 sh = gc.create('Prediksi Nilai Fungsional Tahun 2020')
3 wsResult = gc.open('Prediksi Nilai Fungsional Tahun 2020').s
  heet1

4 set_with_dataframe(wsResult, hasil_pred_akhir2020)

5 sh2 = gc.create('Nilai Karyawan 2019')
6 wsResult2 = gc.open('Nilai Karyawan 2019').sheet1

7 set_with_dataframe(wsResult2, df_pemeriksa2019)

```

Kode Program 4.6 Export Data

BAB 5 HASIL DAN PEMBAHASAN

Pada bagian ini akan membahas mengenai hasil dan analisis dari proses penerapan *Naïve Bayes Classifier* dalam memberikan prediksi nilai pegawai. Selain itu juga akan dibahas uji validasi dengan menggunakan *Accuracy Score* dan *K-Fold Cross Validation*, hasil visualisasi dashboard dan *usability testing*.

5.1 Hasil Uji Validitas

Tahapan uji validitas model perlu dilakukan untuk melihat bagaimana performa dari model *naïve bayes* yang telah dibangun. Semakin besar nilai akurasi yang didapat, maka semakin akurat prediksi yang akan dihasilkan oleh model *naïve bayes* yang dibentuk. Tahap uji validitas model menggunakan dua metode berbeda, yaitu uji validitas menggunakan *Accuracy Score* dan uji validitas menggunakan *K-Fold Cross Validation*.

5.1.1 Accuracy Score

Perhitungan *accuracy score* dilakukan dengan menggunakan fungsi *accuracy_score* yang sudah disediakan oleh library *metrics*. Rasio pembagian data yang digunakan pada *accuracy score* adalah 80% untuk *data training* dan 20% untuk data *test*. Sebelum melakukan tahapan perhitungan *accuracy score*, perlu disiapkan variable yang akan digunakan sebagai parameter, yaitu variable *y_test* dan *y_pred*. Variable *y_test* merupakan variable yang berisi atribut nilai_huruf yang akan diujikan berdasarkan variable *X_test*. Sedangkan variable *y_pred* merupakan variable yang berisi hasil prediksi atribut nilai_huruf berdasarkan variable *X_test*. Untuk hasil dari variable *y_pred* sebagai hasil prediksi dan variable *y_test* sebagai data yang sebenarnya dapat dilihat pada Gambar 5.1 dan Gambar 5.2.

```
[303] y_pred = gnb.predict(X_test)
      y_pred
      array(['A', 'A', 'A', 'D', 'B', 'E', 'D', 'B', 'A', 'D', 'C', 'D', 'A',
            'B', 'D', 'B'], dtype='<U1')
```

Gambar 5.1 Hasil Prediksi Model Naïve Bayes

```
[304] np.array(y_test)
      array(['A', 'A', 'A', 'D', 'B', 'E', 'D', 'B', 'A', 'D', 'C', 'D', 'A',
            'B', 'D', 'A'], dtype=object)
```

Gambar 5.2 Nilai Variable y_test

Dari 16 data prediksi yang ditampilkan pada Gambar 5.1, model *naïve bayes* dapat memprediksi hasil benar sebanyak 15 data dan 1 data yang hasil prediksinya salah. Setelah variable *y_test* dan *y_pred* didefinisikan, kemudian dilakukan pengujian validitas menggunakan fungsi *accuracy_score*. Fungsi *accuracy_score*

digunakan untuk menguji hasil akurasi yang didapat dari prediksi variable `y_pred` terhadap variable `y_test`. Hasil akurasi yang didapatkan sebesar 93,75 %. Hasil akurasi yang didapat dengan menggunakan fungsi `accuracy_score` dapat dilihat pada Gambar 5.3.

```
[305] from sklearn import metrics
      print(metrics.accuracy_score(y_test, y_pred))

0.9375
```

Gambar 5.3 Hasil Akurasi Menggunakan Accuracy Score

5.1.2 K-Fold Cross Validation

Perhitungan *K-Fold Cross Validation* diterapkan pada model *Naïve Bayes* untuk data tahun 2019 dengan jumlah *fold* yang digunakan adalah 5 fold. Sama seperti penggunaan *accuracy score*, rasio pembagian data yang akan digunakan adalah 80% untuk data *training* dan 20% untuk data *testing*. Hasil *cross validation* mendapatkan skor sebesar 82,5 %. Hasil akurasi menurut model *5 fold cross validation* dapat dilihat pada gambar 5.4.

```
#5Fold
from sklearn.model_selection import cross_val_score
scores = cross_val_score(gnb, prediksiX, prediksiY, cv=kf1, scoring='accuracy')
print(scores.mean())

0.825
```

Gambar 5.4 Hasil Akurasi Menggunakan K-Fold Cross Validation

5.2 Prediksi Nilai Pegawai

Hasil prediksi data pegawai tahun 2020 didapat dari penggunaan metode *Naïve Bayes*. Data tahun 2020 digunakan sebagai data uji untuk melihat bagaimana performa model yang telah dibangun dengan menjadikan data tahun 2019 sebagai data *training*. Prediksi dengan metode *Naïve Bayes* dilakukan terhadap atribut data capaian kinerja yaitu Capaian Pencairan SKP, Capaian RD, Capaian Koreksi Rugi SPT, Capaian LHP, Capaian SKPKB Terbit, Capaian Riksus Selesai dan Capaian Produksi Data.

Hasil prediksi nilai pegawai tahun 2020 dapat dilihat dari Tabel 5.1. Dari 52 pegawai yang ada, model *naïve bayes* memprediksi bahwa ada 4 pegawai yang mendapatkan nilai A, 17 pegawai yang mendapatkan nilai B, 7 pegawai yang mendapatkan nilai C, 12 pegawai yang mendapatkan nilai D serta 12 pegawai yang mendapatkan nilai E. Prediksi peringkat pertama diraih oleh pegawai bernama Slamet Priyadi dengan Capaian Pencairan SKP sebesar 120%, Capaian RD sebesar 2.48%, Capaian Koreksi Rugi SPT sebesar 120%, Capaian LHP sebesar 120%, Capaian SKPKB Terbit sebesar 120%, Capaian Riksus selesai sebesar 0% dan Capaian Produksi Data sebesar 63.16%.

Pada hasil prediksi nilai, pegawai dengan nama Ilham Wicaksono dan Karmuji bisa saja mendapatkan nilai A. Namun, model *naïve bayes* memberikan mereka nilai B karena kedua pegawai tersebut memperoleh capaian dengan nilai diatas 100% pada 3 atribut saja. Sedangkan 4 pegawai yang diberikan nilai A oleh model *naïve bayes* memiliki 4 atribut yang bernilai diatas 100%.

Tabel 5.1 Data Hasil Prediksi

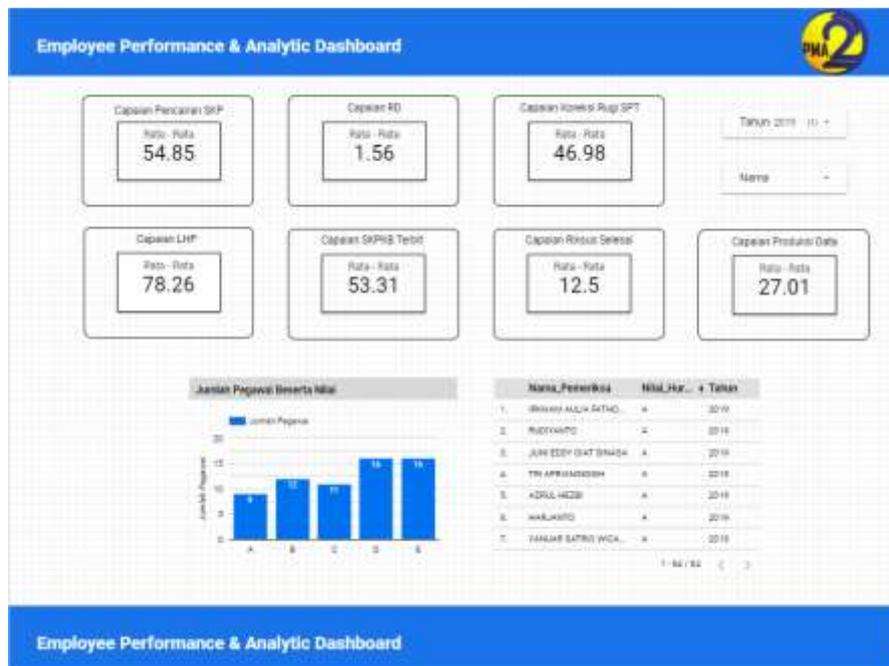
Nama_ Pemeriksa	Pencairan_ SKP_ Capaian	RD_ Capaian	Koreksi_ Rugi_ SPT_ Capaian	LHP_ Capaian	SKPK B_ Terbit_ Capaian	Riksus_ Selesai_ Capaian	Produk si_ Data_ Capaian	Nilai_ Huruf
Slamet Priyadi	120.0	2.48	120.0	120.0	120.0	100.0	63.16	A
Nur Setiawan	120.0	6.11	120.0	120.0	120.0	100.0	61.54	A
Omega Suryadi Sinaga	120.0	3.99	120.0	120.0	120.0	0.0	44.44	A
Egawarman	103.52	7.69	120	120	120	0	8.33	A
Ilham Wicaksono	120	0.26	65.58	120	120	0	15.38	B
Karmuji	120	4.64	60.13	120	120	0	0	B
...
Ivan Rusdian	10.98	0.03	0	81.03	10.68	0	0	E
Wahyu Tri Yanu Wahono	1.7	0.11	0	80.07	1.48	0	88.89	E
Kusnadi	2.87	2.47	0	92.61	46.64	0	0	E
M. Ichsan Riswandi	3.78	0.6	0.0	79.1	5.41	0.0	0.0	E
Triyanto	0.76	0.01	0.0	93.58	2.59	0.0	0.0	E

Edvan Prayudha	0.47	0.4	0.0	86.82	0.47	0.0	0.0	E
----------------	------	-----	-----	-------	------	-----	-----	---

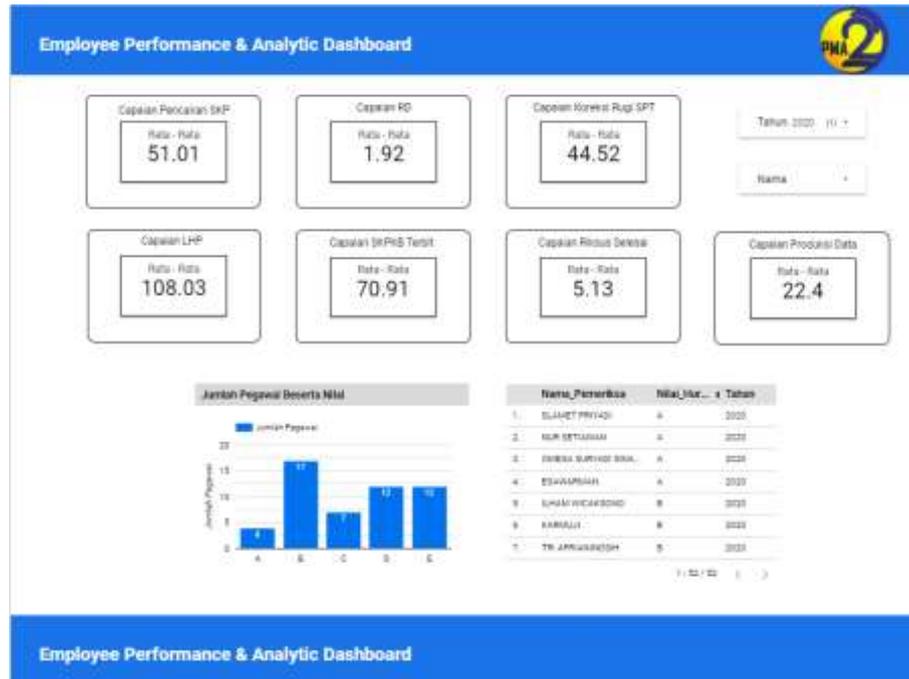
5.3 Hasil Visualisasi Dashboard

Penyusunan *dashboard* perlu dilakukan untuk membantu analisis kinerja pegawai serta menampilkan hasil prediksi dari model *naïve bayes*. Tahap penyusunan *dashboard* dilakukan dengan menggunakan *tools* Google Data Studio serta menggunakan data yang telah diolah menggunakan model *naïve bayes* dan telah di *export* kedalam Google Sheet. Dashboard ini berjudul *Employee Performance & Analytic Dashboard* dan terdiri atas 2 halaman. Halaman pertama menampilkan performa pegawai berdasarkan tahun 2019 dan tahun 2020. Sedangkan halaman kedua menampilkan hasil prediksi nilai pegawai tahun 2020.

Halaman pertama *dashboard* menampilkan rata-rata dari atribut yang ada didalam dataset yang digunakan pada model *naïve bayes*. Atribut-atribut yang ditampilkan adalah dari capaian pencairan SKP, capaian RD, capaian Koreksi Rugi SPT, capaian LHP, capaian SKPKB Terbit, capaian Riksus Selesai, capaian Produksi Data. Selain itu juga terdapat *Bar Chart* untuk menampilkan informasi mengenai jumlah pegawai beserta nilai. Kemudian juga terdapat tabel berisi nama pegawai dan nilai yang diraih. *Dashboard* ini juga memiliki filter untuk berdasarkan tahun dan nama pegawai yang ingin dicari. Gambar 5.5 menunjukkan tampilan *dashboard* data tahun 2019, sedangkan Gambar 5.6 menunjukkan tampilan *dashboard* data tahun 2020.



Gambar 5.5 Dashboard Penilaian Kinerja Pegawai Tahun 2019



Gambar 5.6 Dashboard Penilaian Kinerja Pegawai Tahun 2020

Gambar 5.7 menampilkan halaman kedua *dashboard* yang berisi hasil prediksi nilai pegawai tahun 2020. Hasil prediksi ditampilkan dengan menggunakan tabel yang berisi atribut capaian yang diraih oleh pegawai beserta atribut nilai yang telah diprediksi oleh model *naïve bayes*.

Employee Performance & Analytic Dashboard

Prediksi Nilai Pegawai Tahun 2020

Nama Penerima	Tahun	Pencapaian SKP	PD	Koreksi Rugi S...	LHP	SKPKS	Risiko Seseuai	Produksi Data	Prediksi Nilai
1. NUR BETAYAN	2020	100	6.11	120	120	120	0	97.94	A
2. SLAHET PRIVADI	2020	100	2.68	120	120	120	0	93.16	A
3. DHEGA SURYADI BRAGA	2020	100	9.99	120	120	120	0	84.88	A
4. EDWAPRAN	2020	100.00	7.99	120	120	120	0	83.08	A
5. SUN SUN SUNAWAN	2020	100	7.47	0	95.76	120	0	0	B
6. YAGAR SATNO WICAKSO...	2020	84.49	7.04	0.4	93.08	90.4	0	0	B
7. ED SURPTO	2020	100	0.0	0	120	120	0	5.06	B
8. RIZA MUSLIM	2020	88.24	0.10	0	181.29	99.8	0	12.9	B
9. FERHY KUNIM WICAKSONO	2020	95.46	1.89	64.97	120	120	0	0	B
10. IBINAH ALIA RATHONI	2020	75.70	1.89	0	120	120	30.00	91.82	B
11. FER AHMAD	2020	111.37	0.01	10.72	110.80	120	20.00	16.16	B
12. JAKK SURANA	2020	94.70	1.00	120	116.80	62.96	0	0	B
13. AMES SOBANTO	2020	100	0.68	10.70	115.00	120	0	0	B
14. FARLINDUNGAN BRAGA	2020	77.84	2.11	120	120	82.16	0	0	B
15. SUKMAJI HADIRNO	2020	94.20	1.89	89.27	120	82.11	0	0	B
16. KHRULI	2020	100	4.64	80.70	120	120	0	0	B
17. TRIADI JUMARI	2020	70.71	2.87	0.1	120	71.99	0	17.71	B
18. TEGE APRINZA	2020	85.79	0.19	0	120	87.46	0	28.80	B
19. JAM EDDY GHAT BRAGA	2020	100	2.07	0	120	120	0	18.00	B
20. TRI APRILANDI	2020	88.79	2.00	120	120	120	0	40	B

Gambar 5.7 Dashboard Hasil Prediksi Kinerja Pegawai Tahun 2020

5.4 Hasil *System Usability Scale*

Pengujian *usability* dengan penerapan metode *System Usability Scale* dilakukan untuk mengukur bagaimana kepuasan stakeholder terhadap sistem yang telah dibangun. Pengujian dilakukan terhadap 2 responden, yaitu Kepala Seksi Pemeriksaan KPP PMA Dua serta Tim Pemeriksaan KPP PMA Dua. Hasil pengisian kuesioner *System Usability Scale* dapat dilihat pada Tabel 5.2.

Tabel 5.2 Data Hasil Pengisian *System Usability Scale*

Responden	Pernyataan									
	1	2	3	4	5	6	7	8	9	10
1	5	3	5	4	5	1	5	1	5	5
2	2	1	5	3	5	1	5	1	5	3

Tabel 5.3 Hasil Perhitungan *System Usability Scale*

Responden	Perhitungan Pernyataan										Skor SUS (Total * 2,5)
	1	2	3	4	5	6	7	8	9	10	
1	5-1	5-3	5-1	5-4	5-1	5-1	5-1	5-1	5-1	5-5	77,5
2	2-1	5-1	5-1	5-3	5-1	5-1	5-1	5-1	5-1	5-3	82,5
Rata - Rata											80

Berdasarkan hasil perhitungan skor *System Usability Scale* yang ditunjukkan pada Tabel 5.3, dapat ditarik kesimpulan bahwa *dashboard* yang disusun mendapatkan nilai rata-rata sebesar 80. Dengan nilai tersebut, *dashboard* yang telah dibentuk mendapatkan tingkat penerimaan *acceptable*, *grade scale* bernilai "B", dan *adjective ratings* bernilai "excellent". Dengan nilai yang telah diraih, *dashboard* tersebut dapat diterima dengan baik oleh stakeholder dan tidak perlu dilakukan perbaikan lebih lanjut.

BAB 6 PENUTUP

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan beserta analisis terhadap penelitian yang telah dilakukan, maka ditarik kesimpulan sebagai berikut:

1. Hasil penerapan metode *Naïve Bayes* dapat digunakan untuk memprediksi nilai Fungsional Pemeriksa Pajak dengan cukup baik. Model *naïve bayes* memanfaatkan data pegawai tahun 2019 sebagai *data training*. Berdasarkan uji validitas yang telah dilakukan dengan menggunakan *Accuracy Score*, model *naïve bayes* mendapatkan presentasi akurasi sebesar 93,75%. Sedangkan pada penggunaan *K-Fold Cross Validation*, model *naïve bayes* mendapatkan presentasi akurasi sebesar 82,5%. Model *naïve bayes* juga dimanfaatkan untuk memprediksi kinerja Fungsional Pemeriksa Pajak pada tahun 2020. Dari 52 pegawai, 4 pegawai mendapatkan nilai A, 17 pegawai mendapatkan nilai B, 7 pegawai mendapatkan nilai C, 12 pegawai mendapatkan nilai D serta 12 pegawai mendapatkan nilai E.
2. Untuk membantu pihak KPP PMA Dua dalam mengambil keputusan, maka dibentuklah sebuah *dashboard*. *Dashboard* yang dibentuk memuat informasi yang ditampilkan secara visual agar mampu menyampaikan informasi secara tepat sehingga pengguna dapat mudah menerima informasi tersebut tanpa menimbulkan miskomunikasi. Pengujian menggunakan kuesioner *System Usability Scale* dibutuhkan untuk melihat apa yang pengguna rasakan selama menggunakan *dashboard* yang telah dibangun. Hasil yang diraih memperoleh nilai 80. Hal ini menunjukkan bahwa *dashboard* yang telah dibangun masuk dalam *acceptable* pada kategori *Acceptable*, *excellent* pada kategori *Adjective Ratings* serta B pada kategori *Grade Scale*.

6.2 Saran

Berdasarkan kesimpulan yang telah diraih dari rumusan masalah, ada beberapa saran yang didapatkan agar penelitian berikutnya mendapatkan hasil yang lebih baik.

1. Pada penelitian selanjutnya dapat menggunakan data pada dua tahun kebelakang. Hal ini dilakukan untuk menambahkan jumlah *data training* pada model *naïve bayes*. Sehingga dapat meningkatkan hasil akurasi serta hasil prediksi dengan menggunakan model *naïve bayes*.
2. Penelitian selanjutnya dapat menggunakan nilai yang lebih detail lagi. Sebagai contoh Nilai Pencairan SKP, Nilai RD, Nilai Koreksi Rugi SPT, Nilai LHP, Nilai SKPKB Terbit, Nilai Riksus Selesai dan Nilai Produksi Data. Nilai ini perlu dimanfaatkan agar mendapatkan analisis yang lebih baik.

DAFTAR REFERENSI

- Fadlan, C., Ningsih, S., & Windarto, A. P. (2018). Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 3(1), 1-8.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996, August). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In *KDD (Vol. 96, pp. 82-88)*.
- Firdaus, R. D. (2017). Penerapan Data Mining Untuk Menentukan Klasifikasi Kinerja Pegawai Pada Sistem Informasi Kepegawaian. Tasikmalaya. Universitas Siliwangi.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Informatikalogi.com. Algoritma Naïve Bayes. [online] Tersedia di <<https://informatikalogi.com/algoritma-naive-bayes/>> [Diakses 30 September 2020]
- Larose, D. T. (2005). *An introduction to data mining. Traduction et adaptation de Thierry Vallaud*.
- Oktanisa, I., & Supianto, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 5, 567-576.
- Sari, V. N., Astri, L. Y., & Rasywir, E. (2020). ANALISIS DAN PENERAPAN ALGORITMA NAIVE BAYES UNTUK EVALUASI KINERJA KARYAWAN PADA PT. PELITA WIRA SEJAHTERA. *Jurnal Ilmiah Mahasiswa Teknik Informatika*, 2(1), 53-68.
- Sartika, D., & Sensuse, D. I. (2017). Perbandingan algoritma klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada studi kasus pengambilan keputusan pemilihan pola pakaian. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 3(2), 151-161.
- Siregar, V. M. M. (2018). Sistem Pendukung Keputusan Penentuan Insentif Bulanan Pegawai Dengan Menggunakan Metode Naïve Bayes. *SISTEMASI: Jurnal Sistem Informasi*, 7(2), 78-86.
- Supono, R. A. PERBANDINGAN PENERAPAN DATAMINING DALAM PENILAIAN KINERJA KARYAWAN PADA PERUSAHAAN (Studi kasus pada 3 perusahaan).
- Turban, E., J.E. Aronson dan T.P. Liang. 2005. *Decision Support System and Intelligent Systems - 7th ed*. Pearson Education, Inc. Pearson Education, Inc.
- Dwi Prabantini (penterjemah). 2005. *Sistem Pendukung Keputusan dan Sistem Cerdas*. Penerbit ANDI. Yogyakarta.
- Wibowo, A. P., & Hartati, S. (2016). Sistem Klasifikasi Kinerja Satpam Menggunakan Metode Naïve Bayes Classifier. *INOVTEK Polbeng-Seri Informatika*, 1(2), 192-201.

- Friendly, M. (2008). A brief history of data visualization. In *Handbook of data visualization* (pp. 15-56). Springer, Berlin, Heidelberg.
- Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, 5(Sep), 1089-1105.
- Sutoyo, E., & Fadlurrahman, M. A. Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 6(3), 379-385.
- Defriani, M., Resmi, M. G., & Jaelani, I. (2021). Uji Usability Dengan Metode Cognitive Walkthrough Dan System Usability Scale (SUS) Pada Situs Web STT Wastukencana. *INTECOMS: Journal of Information Technology and Computer Science*, 4(1), 30-39.
- Romli, I., & Putra, B. M. (2021). Evaluasi Penilaian Kinerja Dalam Klasifikasi Data Mining Dengan Metode Naive Bayes. *JURNAL TEKNIK INDUSTRI*, 1(01), 36-45.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.

LAMPIRAN A HASIL WAWANCARA

Berikut hasil wawancara yang telah dilakukan kepada beberapa responden.

Narasumber : Tulus Suparto, S.E., M.Si.
Jabatan : Kepala Seksi Pemeriksaan
Topik Wawancara : Pengumpulan kebutuhan evaluasi kinerja Fungsional Pemeriksa Pajak

P: Pewawancara

N: Narasumber

P: Bagaimana permasalahan dalam mengevaluasi kinerja Fungsional Pemeriksa Pajak?

N: Permasalahan tersebut pasti ada. Salah satunya yaitu kesulitan dalam menggabungkan data kerjanya. Karena datanya perlu diambil dari dua *database* yang berbeda, yaitu SIDJP(Sistem Informasi Direktorat Jenderal Pajak) serta dari ALPP(Aplikasi Laporan Pemeriksaan Pajak). Sehingga memang membutuhkan waktu yang lama untuk menilai kinerja Fungsional Pemeriksaan Pajak.

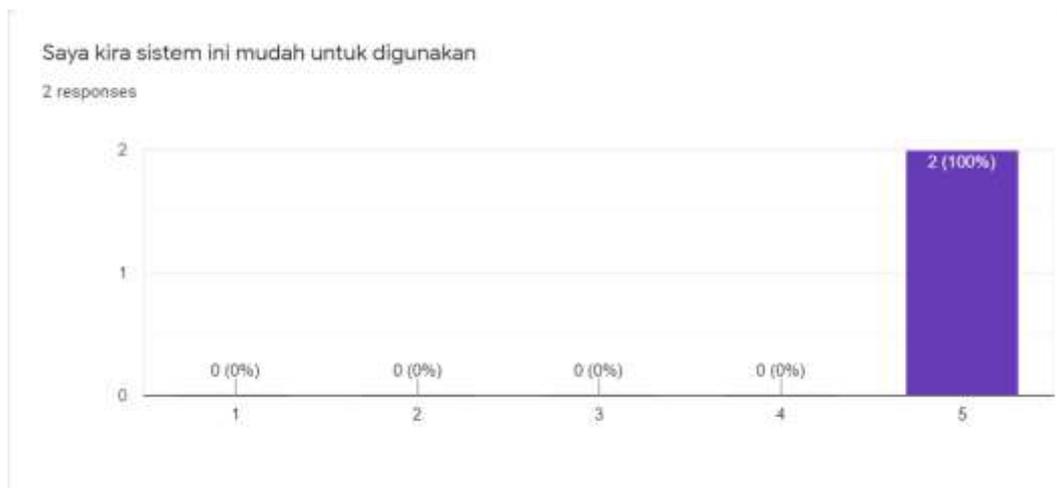
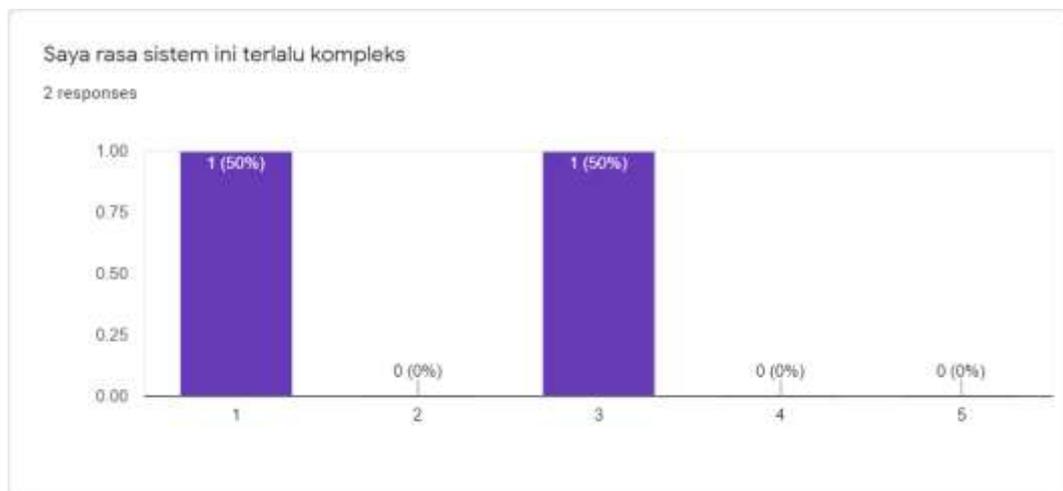
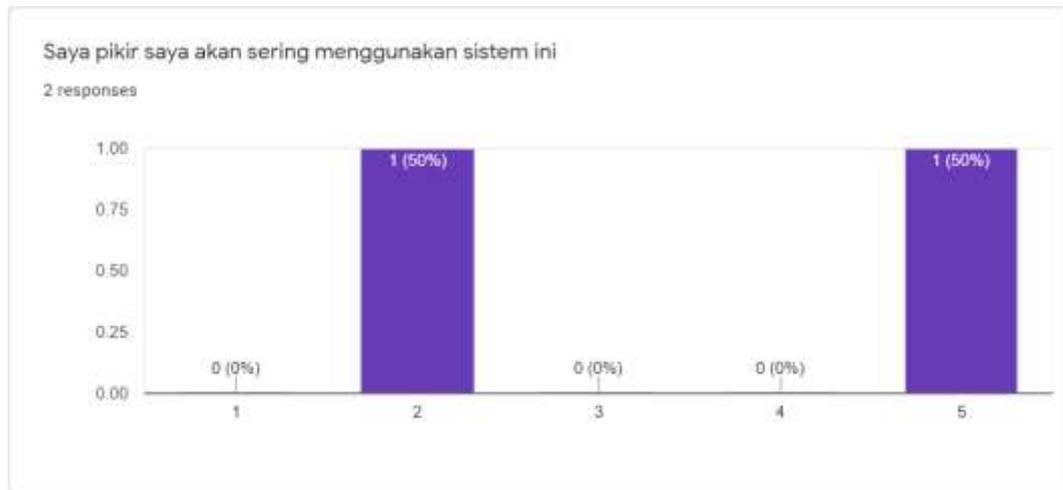
P: Apakah sejauh ini dalam pelaksanaan evaluasi kinerja sudah pernah menggunakan nilai huruf untuk mengukur seberapa baik kinerja pegawainya?

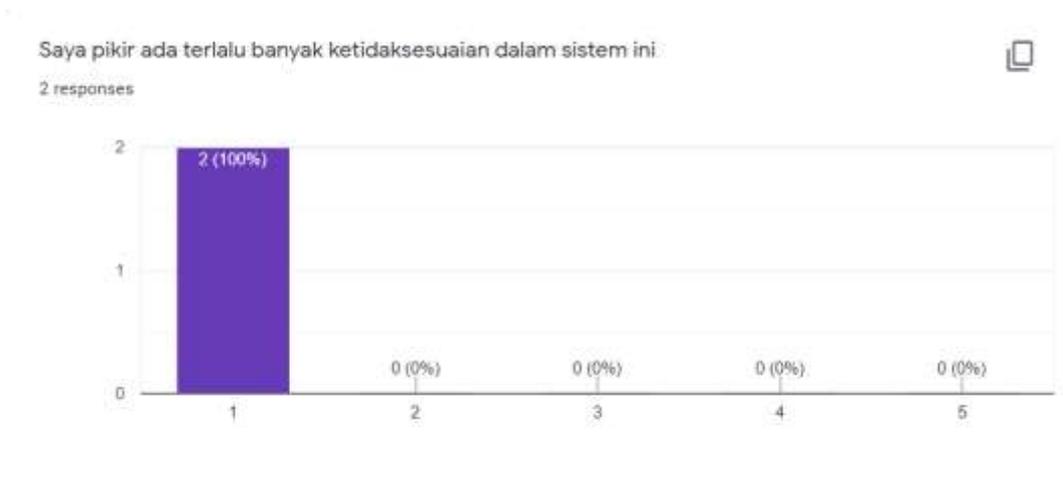
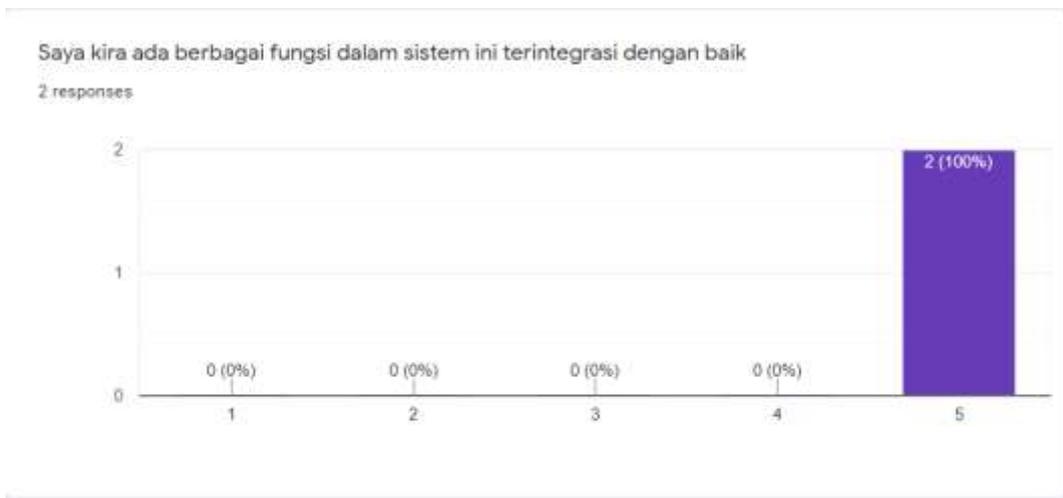
N: Sejauh ini pihak kami hanya menghitung total nilai pegawai saja. Namun kami tertarik untuk menerapkan label nilai huruf bagi pegawai. Sehingga kami sendiri dapat mengelompokkan pegawai-pegawai kedalam nilai huruf mereka.

P: Bagaimana nilai huruf yang seharusnya diterapkan ke pegawai Fungsional Pemeriksa Pajak?

N: Mungkin nilai huruf dapat dikonversi dari total nilai yang didapatkan oleh Fungsional Pemeriksa Pajak. Konversi nilai sendiri dapat diambil seperti nilai A hingga nilai E dengan nilai A diambil dari nilai 80 – 100, nilai B diambil dari nilai 60 – 80, nilai C diambil dari nilai 40 – 60, nilai D diambil dari nilai 20 – 40 dan nilai E diambil dari nilai 0 – 20.

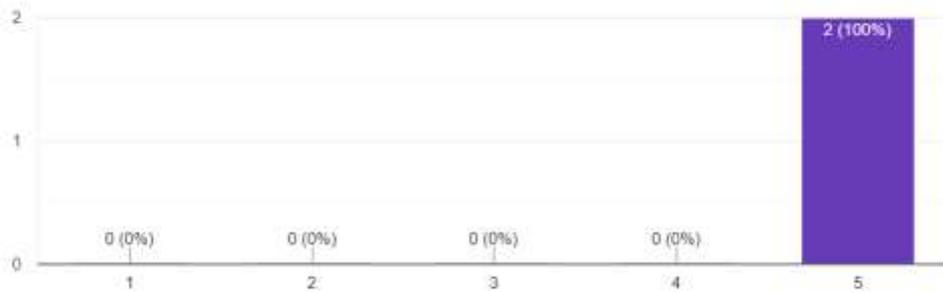
LAMPIRAN B HASIL SYSTEM USABILITY SCALE





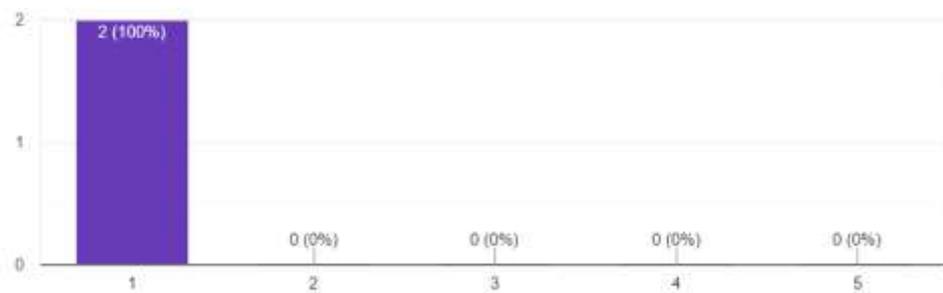
Saya bisa bayangkan bahwa kebanyakan orang akan mudah mempelajari dan menggunakan sistem ini

2 responses



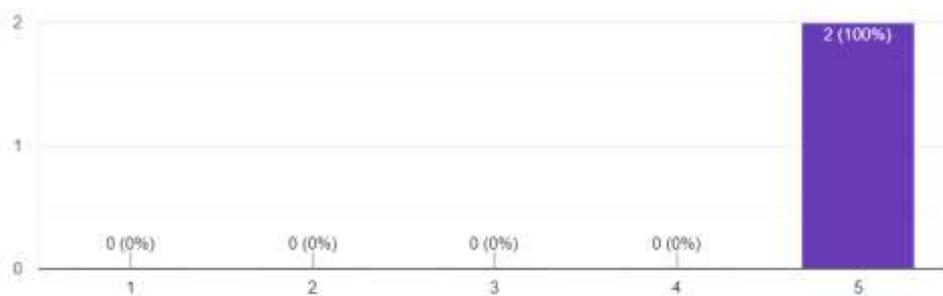
Saya kira sistem ini sangat rumit untuk digunakan

2 responses



Saya merasa percaya diri dalam menggunakan sistem ini

2 responses



Saya perlu belajar banyak hal sebelum saya bisa memulai menggunakan sistem ini

2 responses

