

**PREDIKSI KEBUTUHAN SUKU CADANG MENGGUNAKAN INTEGRASI
CLUSTERING, FORECASTING, DAN ASSOCIATION RULE BERBASIS
MACHINE LEARNING
(STUDI KASUS PT XYZ)**

TUGAS AKHIR

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1
Pada Jurusan Teknik Industri Fakultas Teknologi Industri**



**Nama : Abdul Rahman Rizki Wijaya
No. Mahasiswa : 17 522 113**

**PROGRAM STUDI TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA**

2021

PERNYATAAN KEASLIAN

Dengan ini saya menyatakan bahwa dalam penulisan dan pembuatan penelitian Tugas Akhir ini merupakan hasil dari karya saya sendiri kecuali yang secara tertulis dijelaskan dan diacu pada Tugas Akhir ini. Jika dikemudian hari ditemukan pelanggaran dalam penelitian Tugas Akhir ini, saya siap bertanggung jawab dan menerima risiko terkait permasalahan tersebut.

Yogyakarta, 1 Oktober 2021



Abdul Rahman Rizki Wijaya

LEMBAR PENGESAHAN PEMBIMBING

PREDIKSI KEBUTUHAN SUKU CADANG MENGGUNAKAN INTEGRASI
CLUSTERING, FORECASTING, DAN ASSOCIATION RULE BERBASIS
MACHINE LEARNING
(STUDI KASUS PT XZY)



Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1
Pada Jurusan Teknik Industri Fakultas Teknologi Industri

Disusun Oleh:

Nama : Abdul Rahman Rizki Wijaya

No. Mahasiswa : 17.522.113

Yogyakarta, 1 Oktober 2021

Menyetujui,

Pembimbing

Winda Nur Cahyo, S.T., M.T., Ph.D.

LEMBAR PENGESAHAN DOSEN PENGUJI

**PREDIKSI KEBUTUHAN SUKU CADANG MENGGUNAKAN INTEGRASI
CLUSTERING, FORECASTING, DAN ASSOCIATION RULE BERBASIS
MACHINE LEARNING
(STUDI KASUS PT XZY)**

TUGAS AKHIR

Disusun Oleh:
Nama : Abdul Rahman Rizki Wijaya
No. Mahasiswa : 17 522 113

Telah dipertahankan di depan sidang pengujian sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknik Industri

Yogyakarta, Oktober 2021

Tim Penguji,

Winda Nur Cahyo, S.T., M.T., Ph.D.

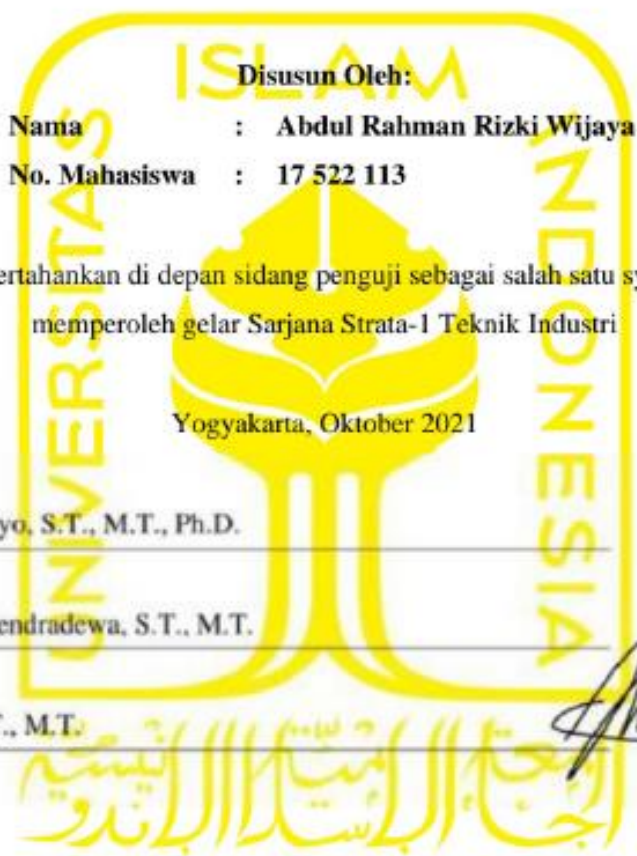
Ketua

Andrie Pasca Hendradewa, S.T., M.T.

Anggota 1

Dian Janari, S.T., M.T.

Anggota 2



Mengetahui,

Ketua Program Studi Teknik Industri

Fakultas Teknologi Industri

Universitas Islam Indonesia



Dr. Fauziq Immawan, S.T., M.M.

HALAMAN PERSEMBAHAN

Tugas Akhir ini saya persembahkan untuk kedua orang tua saya yang telah memberikan dukungan, bimbingan, serta motivasi sehingga saya dapat sampai di titik ini.

Selain itu Tugas Akhir ini juga saya persembahkan untuk bapak/ibu dosen Teknik Industri dan seluruh sahabat serta teman saya selama menempuh pendidikan Strata-1.



MOTTO

"Baginya (manusia) ada malaikat-malaikat yang selalu menjaganya bergiliran, dari depan dan belakangnya. Mereka menjaganya atas perintah Allah. Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum sebelum mereka mengubah keadaan diri mereka sendiri. Dan apabila Allah menghendaki keburukan terhadap suatu kaum, maka tak ada yang dapat menolaknya dan tidak ada pelindung bagi mereka selain Dia." (QS. Ar-Ra'd: 11)

"Dan janganlah kamu iri hati terhadap karunia yang telah dilebihkan Allah kepada sebagian kamu atas sebagian yang lain. (Karena) bagi laki-laki ada bagian dari apa yang mereka usahakan, dan bagi perempuan (pun) ada bagian dari apa yang mereka usahakan. Mohonlah kepada Allah sebagian dari karunia-Nya. Sungguh, Allah Maha Mengetahui segala sesuatu." (An Nisa: 32)

"Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia" (HR. Ahmad, ath-Thabrani, ad-Daruqutni)

KATA PENGANTAR

Assalamu'alaikum Warahmatullaahi Wabarakatuh

Puja dan puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan karunia dan izin-Nya, sehingga penulis dapat menyelesaikan penelitian serta laporan Tugas Akhir. Shalawat serta salam juga penulis aturkan kepada Nabi Muhammad SAW yang telah membawa umat Islam keluar dari zaman kegelapan dan yang kita tunggu syafaatnya di akhir zaman kelak.

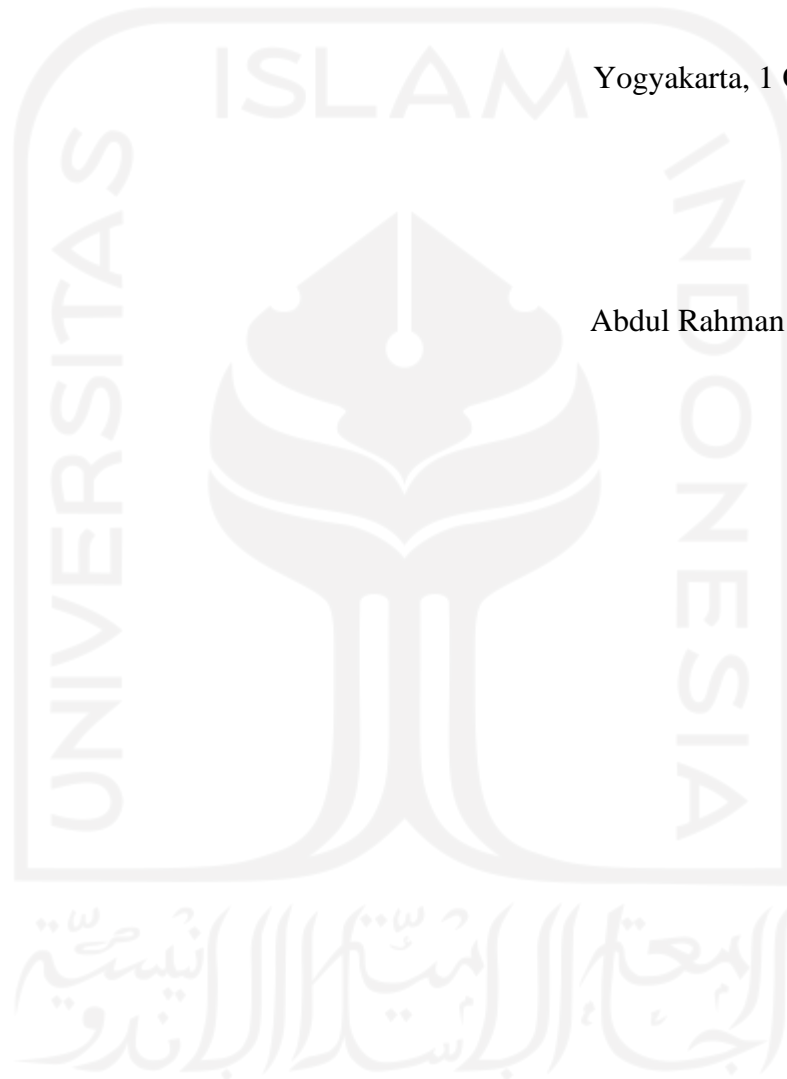
Penelitian dan laporan ini dibuat sebagai salah satu syarat untuk mendapatkan gelar sarjana di Teknik Industri Universitas Islam Indonesia. Di samping itu, penelitian ini menjadi media pembelajaran bagi penulis untuk mengukur pemahaman keilmuan yang didapat selama proses kuliah dan menerapkan ilmu tersebut secara nyata untuk menyelesaikan suatu masalah. Lebih lanjut lagi hasil penelitian dapat menjadi masukan atau alternatif solusi untuk masalah terkait pada perusahaan tempat dilaksanakannya penelitian

Penyelesaian penelitian dan laporan tugas akhir ini tidak lepas dari bantuan, dukungan, dan doa dari banyak pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada pihak-pihak berikut:

1. Kedua Orang Tua yang telah memberikan izin, dukungan, dan doa selama proses penelitian.
2. Bapak Prof. Dr. Ir. Hari Purnomo, M.T. selaku Dekan Fakultas Teknologi Industri.
3. Bapak Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D. selaku Kepala Jurusan Teknik Industri.
4. Bapak Dr. Taufiq Immawan, S.T., M.M. selaku Kapala Program Studi Sarjana Teknik Industri.
5. Bapak Winda Nur Cahyo, S.T., M.T., Ph.D. selaku Dosen Pembimbing yang telah memberikan masukan dan bimbingan.
6. Keluarga Lab ERP yang telah menjadi tempat bercerita dan memberikan dukungan sepenuh hati.
7. Teman-teman Teknik Industri angkatan 2017 yang telah berjuang bersama dan saling mendukung selama masa perkuliahan.
8. Seluruh pihak yang tidak dapat disebutkan satu persatu semoga Allah memberikan karunia-Nya kepada kita semua.

Penulis menyadari bahwa penelitian ini masih memiliki kekurangan. Oleh karena itu, penulis berharap pembaca dapat memberikan kritik dan saran untuk penyempurnaan penelitian yang dilakuakn. Terakhir, penulis berharap laporan ini dapat bermanfaat bagi banyak pihak serta menjadi berkah bagi semua pihak.

Wassalaamu'alaikum Warahmatullaahi Wabarakatuh



Yogyakarta, 1 Oktober 2021

Abdul Rahman Rizki Wijaya

ABSTRAK

PT XYZ merupakan perusahaan yang bergerak dalam industri perminyakan. Dalam operasional setiap harinya PT XYZ melakukan banyak proses bisnis salah satunya proses pengadaan. Proses tersebut memerlukan prediksi kebutuhan barang untuk dilakukan proses pengadaan. Hal tersebutlah yang melatarbelakangi diadakannya penelitian ini. Penelitian ini bertujuan untuk memprediksi tingkat kebutuhan barang sehingga dapat membantu bagian pengadaan untuk melakukan pembelian barang. Metode yang digunakan yaitu K-Means untuk Clustering, Holt Winter untuk peramalan, dan Association Rule untuk mengetahui kombinasi barang yang sering diadakan. Proses clustering menghasilkan 3 kluster yaitu kluster pertama yang memiliki anggota 327 barang dengan nilai rata-rata sebesar 68, kluster kedua yang memiliki anggota 18 barang dengan nilai rata-rata sebesar 1281, dan kluster ketiga yang memiliki anggota 5 barang dengan nilai rata-rata sebesar 2474. Peramalan yang dihasilkan mencakup 3 bulan periode mendatang dengan nilai 427 pada bulan Januari, 472 pada bulan Februari, dan 476 pada bulan Maret. Selain itu juga diketahui barang yang berkemungkinan untuk muncul pada bulan tersebut dengan menggunakan metode peluang serta nilai minimum peluang sebesar 50%. Pada bulan Januari barang yang mungkin untuk muncul sebanyak 10 barang, bulan Februari sebanyak 22 barang, dan bulan Maret sebanyak 20 barang. Aturan yang dihasilkan dari proses association rule berjumlah 24 aturan. Aturan tersebut dihasilkan dengan nilai minimal support sebesar 10% dan minimal confidence sebesar 50%. Integrasikan antar metode juga dilakukan untuk menutupi kekurangan yang ada, sehingga didapatkan hasil yang lebih akurat.

Kata Kunci: Prediksi, K-Means, Holt Winter, Association Rule, Integrasi, Machine Learning

DAFTAR ISI

JUDUL.....	i
PERNYATAAN KEASLIAN	ii
LEMBAR PENGESAHAN PEMBIMBING.....	iii
LEMBAR PENGESAHAN DOSEN PENGUJI.....	iv
HALAMAN PERSEMBAHAN	v
MOTTO	vi
KATA PENGANTAR	vii
ABSTRAK.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
1 BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian	4
1.4. Batasan Masalah.....	5
1.5. Manfaat Penelitian	5
1.6. Sistematika Penulisan Laporan TA	6
2 BAB II KAJIAN PUSTAKA	7
2.1. Kajian Induktif	7
2.2. Kajian Deduktif	11
2.2.1. <i>Knowledge Discovery in Database (KDD)</i>	11
2.2.2. <i>Data Mining</i>	13
2.2.3. <i>Machine Learning</i>	14
2.2.4. <i>Python</i>	15
2.2.5. <i>Clustering</i>	16
2.2.6. <i>K-Means Clustering</i>	17
2.2.7. <i>Forecasting</i>	18
2.2.8. <i>Association Rules (AR)</i>	20
2.2.9. <i>Algoritma Apriori</i>	21
3 BAB III METODE PENELITIAN	22
3.1. Objek Penelitian	22

3.2.	Metode Pengumpulan Data	22
3.3.	Jenis Data	23
3.4.	Alur Penelitian	24
4	BAB IV PEGOLAHAN DATA	27
4.1.	Pengumpulan Data	27
4.1.1.	Data Historis Pengadaan	27
4.1.2.	<i>Preprocessing</i> Data Historis	28
4.2.	Pengolahan Data.....	29
4.2.1.	<i>Clustering</i>	30
4.2.2.	<i>Forecasting</i>	38
4.2.3.	<i>Association Rule</i> (AR)	45
5	BAB V PEMBAHASAN	49
5.1.	<i>Clustering</i>	49
5.2.	<i>Forecasting</i>	52
5.3.	<i>Association Rule</i> (AR).....	54
5.4.	Integrasi.....	56
5.5.	Diskusi.....	62
6	BAB VI KESIMPULAN & SARAN	65
6.1.	Kesimpulan	65
6.2.	Saran.....	66
	DAFTAR PUSTAKA	68
	LAMPIRAN.....	1

DAFTAR TABEL

Tabel 2.1. Perbandingan Antar Jurnal	10
Tabel 4.1. Hasil Perhitungan Peluang Bulan Januari.....	43
Tabel 4.2. Hasil Perhitungan Peluang Bulan Februari.....	44
Tabel 4.3. Hasil Perhitungan Peluang Bulan Maret.....	45
Tabel 4.4. Hasil Aturan AR	48
Tabel 5.1. Hasil Aturan AR	55
Tabel 5.2. Hasil Peluang Bulan Januari	57
Tabel 5.3. Hasil Peluang Bulan Februari	57
Tabel 5.4. Hasil Peluang Bulan Maret	58
Tabel 5.5. Hasil Integrasi Clustering Bulan Januari	59
Tabel 5.6. Hasil Integrasi Clustering Bulan Februari	59
Tabel 5.7. Hasil Clustering Integrasi Bulan Maret	60
Tabel 5.8. Hasil Integrasi AR Bulan Januari	61
Tabel 5.9. Hasil Integrasi AR Bulan Februari	62
Tabel 5.10. Hasil Integrasi AR Bulan Maret	62



DAFTAR GAMBAR

Gambar 2.1. Proses KDD	11
Gambar 3.1. Alur Penelitian	24
Gambar 4.1. Atribut dan Tipe Data	27
Gambar 4.2. Atribut dan Tipe Data Hasil Preprocessing	29
Gambar 4.3. Contoh Data Hasil Preprocessing	29
Gambar 4.4. Syntax Import Modul dan Dataset	30
Gambar 4.5. Contoh 5 Hasil Pengelompokan.....	31
Gambar 4.6. Syntax dan Hasil Perhitungan VIF	32
Gambar 4.7. Syntax dan Hasil Scatter Diagram	33
Gambar 4.8. Syntax dan Hasil Metode Elbow.....	34
Gambar 4.9. Syntax dan Hasil Clustering.....	35
Gambar 4.10. Karakteristik Kluster Pertama.....	36
Gambar 4.11. Karakteristik Kluster Kedua	37
Gambar 4.12. Karakteristik Kluster Ketiga	38
Gambar 4.13. Syntax dan Hasil Transformasi Data	39
Gambar 4.14. Distribusi Data	40
Gambar 4.15. Syntax Pembuatan Model	40
Gambar 4.16. Perbandingan Data Aktual dan Peramalan	41
Gambar 4.17. Tingkat Error Peramalan.....	41
Gambar 4.18. Hasil Peramalan 3 Bulan.....	42
Gambar 4.19. Syntax Perhitungan Peluang	43
Gambar 4.20. Syntax Transformasi Data.....	46
Gambar 4.21. Hasil Transformasi.....	47
Gambar 4.22. Syntax Pembuatan AR	47
Gambar 5.1. Hasil Peramalan	52
Gambar 5.2. Hasil Peramalan	56

BAB I

PENDAHULUAN

1.1. Latar Belakang

Proses bisnis merupakan bagian penting dari sebuah perusahaan. Proses bisnis menjadi salah satu faktor penentu kelancaran, performa, dan keberhasilan perusahaan. Proses bisnis sendiri merupakan serangkaian kegiatan yang saling terhubung untuk mencapai hasil yang diinginkan dalam mendukung berjalannya sebuah perusahaan (Harrington, 1991). Sebuah perusahaan melakukan berbagai macam proses bisnis untuk mendukung berjalannya perusahaan tersebut. Proses ini dapat berupa penjualan, produksi, pengadaan, ataupun proses lainnya. Setiap proses ini perlu untuk diatur dengan baik guna meningkatkan kinerja organisasi secara keseluruhan. Salah satu contoh proses bisnis yang dilakukan oleh perusahaan adalah proses pengadaan spare part. Pengadaan ini dilakukan untuk menjaga tingkat ketersediaan dari spare part tersebut. Terlebih lagi spare part yang digunakan untuk mesin produksi. Jika tingkat ketersediaan tidak mencukupi maka akan mempengaruhi availability (ketersediaan) dan reliability (keandalan) untuk mesin dalam produksi. Jika hal tersebut terus berlanjut akan berdampak buruk untuk perusahaan seperti produktivitas produksi menurun, waktu breakdown mesin semakin panjang, sampai dengan target produksi menjadi terganggu. Oleh karena itu, kebutuhan spare part perlu diprediksi dengan baik agar strategi pengadaan dapat dibuat sebaik mungkin. Di sisi lain, dalam memprediksi kebutuhan spare part mesin terdapat beberapa kesulitan untuk melakukannya. Menurut Bacchetti & Saccani (2012), terdapat beberapa aspek yang menyebabkan permintaan dan manajemen spare part menjadi masalah yang rumit, yaitu tingginya jumlah spare part yang dikelola dan adanya permintaan yang bersifat intermittent ataupun lumpy. Permintaan intermittent adalah permintaan yang terjadi dalam interval tidak menentu dan dengan jumlah yang sangat bervariasi. Sedangkan

permintaan yang bersifat lumpy adalah permintaan dengan waktu yang tidak merata dan jumlah yang dibutuhkan bervariasi. Permintaan bersifat lumpy mengakibatkan kebutuhan investasi persediaan yang meningkat atau waktu respon yang lebih lama daripada yang telah diprediksi. Di sisi lain, proses pengadaan ini menghasilkan data yang merupakan penggambaran dari transaksi atau proses yang dilakukan. Jika data ini dimanfaatkan dengan baik akan dapat memberikan informasi yang bernilai bagi perusahaan dan mendukung manajemen perusahaan, seperti dapat diketahui perkiraan tingkat kebutuhan spare part untuk periode tertentu.

Data merupakan aset yang berharga bagi sebuah perusahaan. Data-data yang dihasilkan oleh perusahaan dapat diolah untuk tujuan tertentu seperti pendukung dalam pengambilan keputusan. Keputusan atau kebijakan yang didukung oleh data akan menghasilkan keputusan yang tepat sasaran atau efektif. Di sisi lain jika data yang dimiliki tidak diolah dengan tepat maka data ini tidak akan memberikan manfaat apa pun. Terdapat juga istilah "*Drowning in Data but Starving for Insight*" yang berarti perusahaan menghasilkan banyak data namun tidak mengolahnya dengan baik sehingga tidak memberikan nilai apa pun (Rizki, et al., 2019). Oleh karena itu data yang dihasilkan sangat penting untuk diolah agar menghasilkan suatu manfaat bagi perusahaan. Selain itu tren perusahaan pada saat ini juga menunjukkan perubahan perusahaan untuk menjadi *data-driven company* ataupun *data-informed company*. *Data-driven* sendiri berarti keputusan yang diambil oleh sebuah perusahaan ditentukan berdasarkan data. Sedangkan, *data-informed* berarti keputusan perusahaan diambil berdasarkan data namun ditambahkan dengan faktor lain seperti pengalaman masa lalu, riset pengguna, atau informasi penting lainnya.

Teknologi menjadi bagian penting untuk banyak aspek kehidupan pada saat ini. Adanya teknologi diharapkan dapat mempermudah pekerjaan yang dilakukan oleh manusia. Selain itu perkembangan teknologi juga semakin pesat dengan adanya inovasi-inovasi baru yang dibuat. Perkembangan teknologi ini terjadi pada banyak aspek termasuk pengolahan data. Penerapan teknologi bagi perusahaan dapat meningkatkan produktivitas dari sebuah perusahaan terlebih lagi dalam pengolahan data. Hal ini karena

semakin besar sebuah perusahaan maka data yang dihasilkan akan semakin besar juga. Sehingga, dengan adanya teknologi pengolahan data akan lebih cepat dibandingkan dengan pengolahan manual serta dapat meminimalkan tingkat kesalahan yang terjadi pada saat proses pengolahan. Terlebih lagi dengan adanya revolusi industri 4.0 akan mengakibatkan data yang dihasilkan sebuah perusahaan semakin besar sehingga dibutuhkan alat bantu untuk mengolah data ini.

Machine learning menjadi salah satu pilihan teknologi untuk pengolahan data. Dengan menggunakan *machine learning* pengolahan data untuk data besar menjadi lebih mudah. Menurut IBM, *machine learning* merupakan cabang aplikasi dari *artificial intelligence* yang berfokus pada pembuatan sistem atau algoritma yang terus belajar berdasarkan data secara terus menerus untuk meningkatkan akurasi. Dalam aplikasi *machine learning*, algoritma dapat digunakan untuk mencari pola atau fitur tertentu pada data berukuran besar untuk membuat keputusan maupun prediksi dari data yang ada. Sehingga, data perusahaan yang cukup banyak dapat diolah dengan baik. Dalam penggunaan *machine learning* dapat menggunakan bahasa pemrograman *Python* karena *Python* menyediakan serta memungkinkan untuk menjalankan modul-modul *machine learning*. Terlebih lagi *Python* merupakan bahasa pemrograman yang populer dalam data science dan merupakan bahasa pemrograman yang cukup mudah dipahami.

PT XYZ merupakan salah satu perusahaan yang ada di Indonesia. Dalam pengoperasian bisnisnya pasti akan dihasilkan data yang banyak dari proses bisnis mereka seperti proses penjualan, pengadaan, dan produksi. Data-data ini harus diolah agar dapat memberikan pengetahuan atau informasi bagi pihak manajemen dan sebagai pendukung dalam pengambilan kebijakan. Salah satu data yang dihasilkan yaitu mengenai pengadaan barang yang dilakukan oleh perusahaan. PT XYZ melakukan banyak proses pengadaan untuk banyak barang, namun dalam penelitian ini hanya akan mencakup pada 1 jenis barang yaitu salah satu suku cadang yang dimiliki oleh PT XYZ.

Dalam penelitian ini akan berfokus pada pengolahan data pengadaan material tahun 2010-2015 dari PT XYZ untuk menemukan informasi berharga yang dapat digunakan oleh pihak manajemen dalam pengambilan keputusan. Informasi yang ingin digali berhubungan dengan kebutuhan dari suku cadang dengan menggunakan data pengadaan PT XYZ. Dengan adanya penelitian ini diharapkan dapat memberikan informasi pendukung untuk pengambilan keputusan dalam proses pengadaan barang tersebut sehingga pengambilan keputusan dapat lebih mudah dan tepat.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah dijabarkan maka rumusan masalah yang diangkat adalah:

- a. Bagaimana hasil kluster serta karakteristik kluster yang terbentuk?
- b. Berapakah prediksi tingkat pengadaan material serta material yang berkemungkinan untuk muncul pada bulan Januari – Maret tahun 2016?
- c. Bagaimana kombinasi material yang sering muncul berdasarkan data transaksi yang dimiliki?
- d. Bagaimana integrasi dari metode yang digunakan?

1.3. Tujuan Penelitian

Tujuan yang ingin dicapai melalui penelitian ini adalah:

- a. Mengetahui hasil kluster serta karakteristik dari kluster yang terbentuk.
- b. Mengetahui tingkat pengadaan material dan material yang berkemungkinan untuk muncul pada bulan Januari – Maret tahun 2016.
- c. Mengetahui kombinasi material yang sering muncul berdasarkan data transaksi yang dimiliki.
- d. Mengetahui integrasi dari metode yang digunakan.

1.4. Batasan Masalah

Dalam melakukan penelitian ini akan ditentukan batasan penelitian sehingga lingkup penelitian dapat tepat sasaran dan tidak meluas ke permasalahan lainnya. Berikut ini merupakan batasan masalah yang ditetapkan:

- a. Penelitian menggunakan data pengadaan material PT XYZ dari tahun 2010-2015.
- b. Penelitian hanya menggunakan satu jenis material.
- c. Permasalahan yang diangkat hanya terkait permintaan suatu material.

1.5. Manfaat Penelitian

Manfaat penelitian dapat dirasakan oleh beberapa pihak sebagai berikut:

1.5.1. Bagi Perusahaan

- a. Perusahaan mendapatkan rekomendasi terkait kebutuhan barang.
- b. Perusahaan dapat mengetahui karakteristik data yang mereka miliki.
- c. Pengambilan keputusan menjadi lebih terbantu karena memiliki gambaran terkait data yang dimiliki.

1.5.2. Bagi Mahasiswa

- a. Melatih keahlian analisis mahasiswa terhadap suatu masalah.
- b. Mahasiswa dapat menerapkan secara langsung pengetahuan yang didapat pada saat perkuliahan untuk menyelesaikan suatu permasalahan.
- c. Mahasiswa mengetahui proses dari suatu bisnis lebih mendalam secara langsung.

1.6. Sistematika Penulisan Laporan TA

BAB I PENDAHULUAN

Pendahuluan akan berisi mengenai latar belakang, rumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Kajian Pustaka akan berisi mengenai kajian induktif dan deduktif terkait penelitian dan teori yang relevan. Bab ini akan menjadi dasar teori yang digunakan dalam penelitian.

BAB III METODE PENELITIAN

Penjelasan mengenai metode yang digunakan untuk penelitian akan dijelaskan pada bagian ini. Selain itu juga akan dijelaskan mengenai hal lainnya terkait penelitian seperti objek penelitian, data yang digunakan, serta alur penelitian yang dilakukan.

BAB IV PENGOLAHAN DATA

Bagian ini akan berisi mengenai proses pengolahan serta hasil menggunakan metode yang telah ditentukan sebelumnya.

BAB V PEMBAHASAN

Pembahasan berisikan mengenai analisis dan penjelasan mengenai hasil dari pengolahan data yang sebelumnya dilakukan.

BAB VI KESIMPULAN DAN SARAN

Bagian ini berisikan mengenai kesimpulan dari penelitian yang telah dilakukan. Pada bagian ini juga terdapat saran terkait penelitian yang diangkat.

BAB II

KAJIAN PUSTAKA

2.1. Kajian Induktif

Kajian induktif akan membahas mengenai penelitian-penelitian terdahulu yang memiliki keterkaitan topik. Keterkaitan ini dapat berupa metode penelitian yang digunakan, objek penelitian, atau hal lainnya. Sehingga, dapat diketahui kesamaan dan perbedaan dengan penelitian yang telah dilakukan. Hal tersebut dilakukan untuk mengetahui posisi atau letak dari penelitian yang dilakukan. Selain itu kajian ini juga digunakan sebagai referensi terkait metode, aplikasi, maupun hal lainnya dalam melakukan penelitian.

Nasari et al. (2016) melakukan penelitian mengenai pengelompokan penyebaran diare di Kabupaten Langkat. Penelitian ini bertujuan untuk mencari tahu kawasan penyebaran diare agar tindakan preventif dapat dilakukan. Oleh karena itu, penelitian menggunakan metode *clustering* dengan algoritma *K-Means*. Dengan menggunakan metode tersebut akan dapat memungkinkan untuk mengelompokkan daerah-daerah pada Kabupaten Langkat ke dalam kluster tertentu berdasarkan kedekatan nilai. Data yang digunakan merupakan jumlah penderita diare pada setiap Kecamatan di Kabupaten Langkat. Hasil dari penelitian ini terbentuk dua kluster. Kluster pertama merupakan kluster yang memiliki penderita diare tingkat menengah sedangkan kluster kedua menjadi pusat penderita diare pada Kabupaten Langkat.

Penelitian terkait *clustering* juga dilakukan oleh Rofiqo et al. (2018). Penelitian tersebut menggunakan algoritma *K-Means* untuk mencari kelompok terkait dengan keluhan Kesehatan di Indonesia. Data yang digunakan merupakan data dari Badan Pusat

Statistik (BPS) terkait penduduk provinsi yang memiliki keluhan kesehatan. Data ini terdiri dari 34 provinsi serta memiliki rentang waktu antara 2013-2017. Penelitian ini menghasilkan 3 kluster yaitu kluster dengan tingkat keluhan kesehatan tinggi, sedang, dan rendah. Kluster dengan tingkat keluhan kesehatan tinggi terdiri dari 7 provinsi, kluster dengan tingkat keluhan kesehatan sedang terdiri dari 18 provinsi, dan kluster dengan tingkat keluhan kesehatan rendah terdiri dari 9 provinsi.

Gustientiedina et al. (2019) melakukan penelitian *clustering* terkait data obat-obatan RSUD Pekanbaru. Tujuan dari penelitian ini adalah pengelompokan data obat-obatan RSUD Pekanbaru untuk dapat digunakan sebagai referensi pengambilan keputusan dalam hal perencanaan dan pengendalian pasokan medis. Algoritma *clustering* yang digunakan adalah *K-Means* karena algoritma tersebut mengelompokkan objek dengan karakteristik yang sama ke dalam satu kluster dan objek dengan karakteristik yang berbeda ke dalam kluster yang berbeda. Dalam penelitian ini dihasilkan 3 kluster yaitu kluster pemakaian obat sedikit, sedang, dan tinggi. Kluster sedikit memiliki jumlah pemakaian kurang dari 18.000, kluster sedang memiliki jumlah pemakaian antara 18.000-70.000, dan kluster tinggi memiliki jumlah pemakaian di atas 70.000 setiap tahunnya.

Gamberini et al. (2010) melakukan penelitian terkait permintaan yang sporadis dan tidak teratur. Penelitian ini bertujuan untuk membandingkan kinerja metode peramalan (S)ARIMA dan Holt Winter untuk memprediksi pesanan pelanggan yang tidak teratur. Adanya penelitian ini dapat membantu perusahaan baru yang cenderung memiliki tipe permintaan pelanggan yang tidak teratur karena belum terlalu dikenal. Pengukuran kinerja akan menggunakan nilai Mean Absolute Deviation (MAD) dan Mean Square Error (MSE). Kesimpulan dari penelitian ini bahwa jika data tidak memiliki trend yang konsisten, kedua metode memberikan kinerja yang mirip atau tidak menunjukkan perbedaan yang signifikan. Namun, jika data memiliki trend yang konsisten, metode (S)ARIMA menunjukkan hasil kinerja yang lebih baik.

Feng et al. (2019) melakukan penelitian untuk memprediksi konsumsi spare part yang memiliki sifat seasonal pada bidang penerbangan. Metode prediksi yang digunakan yaitu Holt Winters exponential smooting. Berdasarkan penelitian ini diketahui metode ini menghasilkan nilai error yang kecil serta perhitungan yang cukup sederhana. Hasil ini membantu pihak manajemen dalam pengambilan keputusan dalam hal spare part untuk memberikan pelayanan yang lebih baik.

Aprianti et al. (2017) melakukan penelitian dengan menggunakan metode *association rule*. Penelitian tersebut terkait dengan indikator-indikator kemiskinan serta bagaimana indikator-indikator tersebut saling berkaitan. *Dataset* yang digunakan terdiri dari 46 atribut yang berasal dari BPS Kabupaten Tanah Laut dan BPS Provinsi Kalimantan Selatan tahun 2010-2014. Data tersebut akan diolah dengan menggunakan algoritma apriori serta ditentukan untuk nilai minimum *support* sebesar 30% dan minimum *confidence* sebesar 80%. *Rule* atau aturan yang dihasilkan setelah dilakukan pengolahan terbentuk 4614 kombinasi atribut. Hasil tersebut masih terlalu luas untuk dapat dilakukan pengambilan kesimpulan oleh karena itu peneliti memberikan saran melakukan *principal component analysis* pada saat *preprocessing* untuk menghasilkan aturan yang lebih terperinci.

Dalam penelitian yang dilakukan oleh Rerung (2018) digunakan metode *association rule* dengan algoritma apriori untuk mengetahui kombinasi item yang sering muncul. Tujuan dari penelitian tersebut adalah mengetahui item yang sering terjual dan kombinasinya untuk pembuatan strategi promosi produk. Data yang digunakan berasal dari Distro Nasional. Selain itu dalam penelitian ini juga dikembangkan aplikasi penunjang strategi promosi dengan menerapkan metode *association rule* yang telah dibuat. Hasil dari penelitian ini adalah aplikasi yang menunjukkan hasil *association rule* atau kombinasi item yang sering untuk muncul dengan nilai *support* dan *confidence* yang dapat disesuaikan.

Penggunaan metode *association rule* digunakan oleh Fauzy et al. (2016) untuk memprediksi cuaca yang akan terjadi. Algoritma apriori dipilih untuk pengolahan *association rule*. *Dataset* yang digunakan dalam penelitian ini adalah data klimatologi BMKG stasiun geofisika kelas 1 Bandung. Hasil dari penelitian ini merupakan aturan-aturan asosiasi yang dapat digunakan untuk memperkirakan apakah akan turun hujan atau tidak pada satu hari ke depan. Namun, penelitian ini memiliki catatan seperti semakin tinggi nilai minimum *support* dan nilai minimum *confidence* akan mengakibatkan jumlah *frequent itemset* dan *rule* yang terbentuk semakin kecil serta menurunnya tingkat akurasi. Di sisi lain setiap aturan yang dihasilkan memiliki nilai *lift ratio* lebih besar daripada 1 sehingga dapat digunakan sebagai acuan dalam memprediksi hujan.

Tabel 2.1. Perbandingan Antar Jurnal

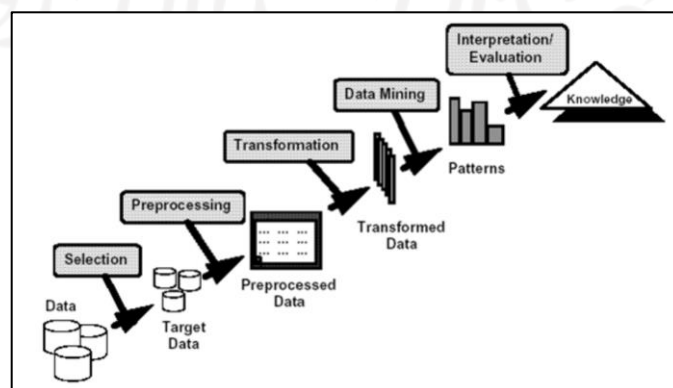
No	Author	Tahun	Metode					
			K-Means	Holt Winters	Apriori	Clustering	Forecasting	Association Rule
1	Nasari, Jhony, & Sianturi	2016	v			v		
2	Rofiqo, Windarto, & Hartama	2018	v			v		
3	Gustientiedina, Adiya, & Desnelita	2019	v			v		
4	Rita Gamberini, Francesco Lolli, Bianca Rimini, & Fabio Sgarbossa	2010		v			v	
5	Zhang Feng, Chen Bing, Wang Shouquan, Cui Gaolun, & Guo Weibo	2019		v			v	
6	Aprianti, Hafizd, & Rizani	2017			v			v
7	Rerung	2018			v			v
8	Fauzy, Saleh W, & Asror	2016			v			v
9	Rizki	2021	v	v	v	v	v	v

2.2. Kajian Deduktif

Kajian deduktif merupakan landasan teori yang digunakan dalam memecahkan masalah penelitian. Kajian ini digunakan untuk menerangkan konsep dari metode dan hal lainnya yang berkaitan dengan penelitian. Referensi yang digunakan untuk melengkapi kajian deduktif berupa jurnal, buku, artikel ilmiah, serta sumber lainnya yang relevan dan terpercaya. Dengan adanya kajian ini dapat menjelaskan dan memberikan gambaran terhadap landasan dari penelitian yang dilakukan.

2.2.1. *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database diartikan sebagai proses memperoleh informasi potensial, implisit, dan tidak dikenal dari sekelompok data yang prosesnya melibatkan data *mining*, kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami (Andayani, 2007). Menurut Zanuardi & Suprayitno (2018), KDD merupakan proses yang terstruktur untuk memperoleh informasi yang benar, baru, bermanfaat, serta menemukan pola dari data yang berukuran besar dan rumit. Dalam melakukan KDD terdapat proses yang perlu untuk dijalankan yaitu data *selection*, *preprocessing*, *transformation*, data *mining*, dan *evaluation/interpretation* (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Gambar 2.1. menunjukkan alur proses KDD.



Gambar 2.1. Proses KDD

Penjelasan terkait proses tersebut adalah sebagai berikut:

1. *Data Selection*

Seleksi atau pemilihan data perlu dilakukan sebelum melakukan proses penggalian informasi dalam KDD. Hal ini dilakukan untuk menyesuaikan kebutuhan data pada pengolahan data. Data yang telah terseleksi akan disimpan pada tempat yang terpisah dari basis data operasional.

2. *Preprocessing*

Sebelum proses data *mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Selain itu juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data *mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

Data *mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data *mining* sangat bervariasi. Pemilihan metode dan algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Evaluation/Interpretation*

Pola informasi yang dihasilkan dari proses data *mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.2.2. Data Mining

Data *mining* atau dapat juga disebut dengan penambangan data merupakan proses menemukan informasi baru dengan mencari pola atau aturan tertentu dalam sekelompok data yang berukuran besar (Davies & Beynon, 2004). Data *mining* juga dapat diartikan sebagai serangkaian proses untuk menghasilkan nilai tambah berupa pengetahuan baru yang selama ini belum diketahui jika menggunakan metode manual dari suatu kumpulan data (Pramudiono, 2003).

Data *mining* adalah proses untuk menemukan pola dari sejumlah besar data, data tersebut dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. Data *mining* juga masih berkaitan dengan ilmu-ilmu lainnya seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu data *mining* juga didukung oleh ilmu seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* (Han & Kamber, 2006). Proses data *mining* dapat dilakukan dengan otomatis atau *semiotomatis* serta pola yang dihasilkan harus memiliki arti yang bermanfaat.

Data *mining* memiliki karakteristik sebagai berikut:

- Data *mining* berhubungan dengan menemukan informasi dan pola data tertentu yang belum diketahui sebelumnya.
- Data *mining* menggunakan data yang berukuran besar. Hal ini dikarenakan data tersebut digunakan untuk meningkatkan akurasi hasil.
- Data *mining* digunakan untuk membuat keputusan yang penting seperti dalam pembuatan strategi perusahaan (Davies & Beynon, 2004).

Berdasarkan pengertian di atas dapat ditarik kesimpulan bahwa data *mining* merupakan cara untuk menggali informasi berharga yang tersembunyi dari sebuah *dataset* yang besar sehingga ditemukan pola data yang berharga. Beberapa metode yang sering

disebutkan dalam literatur data *mining* antara lain *clustering*, *classification*, *association rules*, *neural network*, *genetic algorithm*, dan lain-lain (Pramudiono, 2003).

2.2.3. *Machine Learning*

Machine learning merupakan serangkaian metode yang digunakan untuk membantu memproses dan memprediksi data yang memiliki ukuran besar dengan mempresentasikan data-data tersebut dengan algoritma tertentu (Danukusumo, 2017). Istilah *machine learning* sendiri pertama kali diusulkan oleh Arthur Samuel pada tahun 1959. Menurut Arthur Samuel, *machine learning* merupakan salah satu bidang komputer yang memberikan kemampuan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas. Menurut Mohri et al. (2012), *machine learning* dapat diartikan sebagai metode komputasi berdasarkan pembelajaran untuk meningkatkan kinerja atau melakukan prediksi yang akurat. Arti dari pengalaman sendiri adalah informasi yang telah tersedia atau informasi masa lalu yang akan menjadi data pembelajaran. *Machine learning* juga dapat diartikan sebagai sebuah metode yang memungkinkan untuk komputer memiliki kemampuan mempelajari dan melakukan sebuah pekerjaan secara otomatis. Proses *machine learning* dilakukan melalui algoritma tertentu sehingga pekerjaan yang dijalankan oleh komputer dapat dilakukan dengan otomatis (Nugraha & Hairani, 2018).

Proses *machine learning* dilakukan dengan menggunakan fase *training* dan fase *application*. Fase *training* yaitu proses pemodelan data menggunakan algoritma tertentu oleh sistem melalui data *training*. Fase *application* adalah proses menggunakan model yang telah dibuat sebelumnya pada fase *training* untuk digunakan menghasilkan keputusan tertentu dengan menggunakan testing data. *Machine learning* dapat dijalankan dengan dua acara yaitu *supervised learning* dan *unsupervised learning*. *Supervised learning* adalah pemrosesan sampel data x akan diproses sedemikian rupa sehingga menghasilkan *output* yang sesuai dengan hasil akhir y . Contoh *supervised learning* adalah metode klasifikasi. *Unsupervised learning* adalah pemrosesan sampel data dilakukan

tanpa mewajibkan hasil akhir memiliki bentuk yang sesuai dengan dengan bentuk tertentu. Contoh dari *unsupervised learning* adalah metode *clustering*. Dengan kata lain skenario *machine learning* adalah sebagai berikut:

1. *Supervised Learning*

Penggunaan skenario *supervised learning*, pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.

2. *Unsupervised Learning*

Penggunaan skenario *unsupervised learning*, pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokkan data berdasarkan karakteristik-karakteristik yang ditemui.

2.2.4. *Python*

Python merupakan salah satu bahasa pemrograman yang dikembangkan oleh Guido van Rossum pada tahun 1990 di CWI, Amsterdam. Menurut Pane & Saputra (2020), *Python* adalah bahasa pemrograman yang dapat digunakan untuk memberikan tingkat keterbacaan suatu kode untuk dapat melakukan sejumlah perintah multiguna secara langsung dengan berorientasi pada objek serta menggunakan semantik dinamis. Dalam *Python* terdapat suatu *library* yang menyediakan fungsi analisis data, *machine learning*, data *preprocessing tools*, serta visualisasi data. Oleh karena itu *Python* menjadi bahasa pemrograman yang populer untuk digunakan dalam data *science* dan analisis. Selain itu *Python* juga dapat dijalankan pada berbagai macam sistem operasi seperti Windows, Mac, dan Linux. *Python* juga memiliki *interface* yang mudah dipahami dan dipelajari walaupun tergolong dalam bahasa pemrograman tingkat tinggi.

2.2.5. *Clustering*

Clustering adalah proses pengelompokan sebuah populasi ke dalam kelompok yang lebih kecil dan lebih homogen (Linoff & Berry, 2011). Menurut Everitt et al. (2011), metode analisis *cluster* berhubungan dengan eksplorasi dan menilai kumpulan data agar dapat diringkas menjadi sejumlah kelompok yang *relative* sedikit dengan ketentuan data yang memiliki kemiripan akan menjadi satu kelompok dan data yang berbeda akan masuk ke dalam kelompok lainnya. Dalam hal pemahaman data, *cluster* mampu mengelompokkan data dan analisis terkait *cluster* merupakan bidang ilmu yang secara teknis dapat melakukan pengelompokan data secara otomatis. Contoh *clustering* sederhana yaitu pengelompokan gambar bangunan, tumbuhan, dan hewan oleh anak-anak (Hall, Lynch, Michael, & Mitchell, 2007).

Clustering dapat dibagi menjadi 2 tipe yaitu *soft clustering* dan *hard clustering*. Dalam *soft clustering* memungkinkan untuk sebuah data tergabung dalam lebih dari satu kluster dengan menambahkan faktor probabilitas data tersebut. Dalam *hard clustering*, data hanya akan tergabung ke dalam 1 kluster atau tidak memungkinkan untuk tergabung ke dalam kluster lain. Selain itu karena *clustering* merupakan proses yang subjektif maka banyak cara atau metode yang dapat digunakan untuk melakukan *clustering*. Setiap metode mengikuti aturan tersendiri untuk mendefinisikan kesamaan dari setiap data. Faktanya terdapat lebih dari 100 algoritma *clustering* yang diketahui saat ini, namun sedikit algoritma yang populer digunakan. Algoritma tersebut mengadaptasi model tertentu dalam alur prosesnya. Model yang populer untuk digunakan adalah sebagai berikut:

a. *Connectivity models*

Seperti namanya, model ini didasarkan pada gagasan bahwa titik data yang lebih dekat dalam ruang data menunjukkan lebih banyak kesamaan satu sama lain daripada titik data yang terletak lebih jauh. Model-model ini dapat mengikuti dua pendekatan. Pada pendekatan pertama, dimulai dengan mengklasifikasikan semua titik data ke dalam kelompok yang terpisah & kemudian menggabungkannya saat jarak berkurang. Pada pendekatan kedua,

semua titik data dikelompokkan sebagai satu *cluster* dan kemudian dipartisi seiring bertambahnya jarak. Juga, pilihan fungsi jarak bersifat subjektif. Model-model ini sangat mudah untuk diinterpretasikan tetapi tidak memiliki kapabilitas untuk menangani kumpulan data yang besar. Contoh model tersebut adalah algoritma *clustering* hierarkis dan variannya.

b. *Centroids models*

Model ini adalah algoritma pengelompokan berulang di mana gagasan kesamaan diturunkan oleh kedekatan titik data ke pusat *cluster*. Algoritma *clustering K-Means* adalah algoritma populer yang termasuk dalam kategori ini. Dalam model ini, jumlah *cluster* yang diperlukan pada bagian akhir harus disebutkan sebelumnya atau pada awal pemrosesan. Oleh karena itu penting untuk mengetahui karakteristik dari *dataset* yang dimiliki agar jumlah kluster yang ditentukan dapat optimal. Model-model ini berjalan secara berulang untuk menemukan *local optima*.

c. *Distribution models*

Model pengelompokan ini didasarkan pada gagasan tentang seberapa besar kemungkinan bahwa semua titik data dalam *cluster* memiliki distribusi yang sama (Misalnya: Normal, Gaussian). Namun, model-model ini sering mengalami *overfitting*. Contoh populer dari model ini adalah algoritma Ekspektasi-maksimalisasi yang menggunakan distribusi normal *multivariat*.

d. *Density models*

Model-model ini mencari ruang data untuk area dengan kepadatan bervariasi dari titik data dalam ruang data. Model ini mengisolasi berbagai wilayah kepadatan yang berbeda dan menetapkan titik data dalam wilayah ini di *cluster* yang sama. Contoh model densitas yang populer adalah DBSCAN dan OPTICS.

2.2.6. *K-Means Clustering*

Menurut MacQueen (1967), *K-Means clustering* adalah algoritma *unsupervised machine learning* yang paling banyak digunakan untuk memartisi kumpulan data yang diberikan

ke dalam satu set k grup (k *cluster*), di mana k mewakili jumlah grup yang ditentukan sebelumnya oleh analisis. Algoritma ini mengklasifikasikan objek dalam beberapa kelompok (yaitu, *cluster*), sehingga objek dalam *cluster* yang sama adalah semirip mungkin (kesamaan intra-kelas yang tinggi), sedangkan objek dari *cluster* yang berbeda tidak serupa mungkin (*inter*-kelas rendah). Dalam pengelompokan *K-Means*, setiap *cluster* diwakili oleh pusatnya (*centroid*) yang sesuai dengan rata-rata poin yang ditugaskan ke *cluster*. Menurut Ningrat et al. (2016), langkah-langkah melakukan proses *clustering* dengan menggunakan algoritma *K-Means* adalah sebagai berikut:

1. Tentukan jumlah kluster yang diinginkan.
2. Pilih nilai acak untuk menjadi nilai pusat kluster awal (*centroid*) sebanyak jumlah kluster.
3. Hitung jarak setiap *input* data pada setiap *centroid* menggunakan rumus *Euclidean Distance* sampai menemukan jarak terdekat untuk setiap data dengan *centroid*.
4. Kelompokkan setiap data berdasarkan kedekatan dengan *centroid* (jarak terkecil).
5. Perbaharui nilai *centroid*.
6. Lakukan langkah ke 2 - 5 sampai anggota setiap kluster tidak berubah.
7. Jika langkah 6 telah terpenuhi, maka nilai rata-rata pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

2.2.7. *Forecasting*

Forecasting atau peramalan merupakan proses memprediksi kejadian yang akan datang. Menurut Buffa & Sarin (1996), peramalan menggunakan metode-metode statistik dalam bentuk gambaran masa depan dengan mengolah data-data historis. Pendapat lain diungkapkan oleh Biegel (1999) dengan mendefinisikan peramalan sebagai kegiatan memperkirakan tingkat permintaan produk yang diharapkan atau beberapa produk dalam periode waktu tertentu di masa depan.

Berdasarkan perencanaan operasi di masa depan peramalan dapat dibagi menjadi beberapa tipe. Heizer & Render (2005) membagi peramalan menjadi 3 tipe sebagai berikut:

1. Peramalan Ekonomi

Peramalan ini menjelaskan siklus bisnis dengan memprediksi tingkat inflasi, ketersediaan uang, dana yang dibutuhkan untuk membangun perumahan, serta indikator perencanaan lainnya.

2. Peramalan Teknologi

Peramalan ini memperhatikan tingkat kemajuan teknologi untuk dapat menghasilkan produk baru yang menarik.

3. Peramalan Permintaan

Peramalan ini memprediksi proyeksi permintaan untuk suatu produk atau jasa dari sebuah perusahaan.

Pengelompokan lainnya yang digunakan untuk membedakan peramalan adalah berdasarkan horizon waktu yang dicakupnya. Menurut Taylor (2004), peramalan berdasarkan horizon waktu dapat dibagi menjadi 3 kelompok sebagai berikut:

1. Peramalan Jangka Panjang

Umumnya peramalan ini digunakan untuk memprediksi 2 sampai 10 tahun yang akan datang.

2. Peramalan Jangka Menengah

Umumnya peramalan ini digunakan untuk memprediksi 1 sampai 24 bulan yang akan datang.

3. Peramalan Jangka Pendek

Umumnya peramalan ini digunakan untuk memprediksi 1 sampai 5 minggu yang akan datang.

2.2.8. Association Rules (AR)

Association Rule merupakan metode data *mining* untuk menemukan aturan asosiatif dari suatu kombinasi item atau suatu pekerjaan untuk mengetahui atribut yang muncul bersamaan. *Association rule* biasanya menggunakan aturan “If” dan “Then”. Misalnya “if A then B and C”, hal ini menunjukkan jika A muncul maka B dan C juga akan muncul. Dalam menentukan *association rule* perlu ditentukan *support* dan *confidence* untuk membatasi apakah *rule* tersebut *interesting* atau tidak. Contoh *association rule* ini adalah analisis pembelian pada pasar swalayan untuk mengetahui tingkat kemungkinan seseorang membeli dua atau lebih barang (contoh roti dan susu) dalam sekali pembelian. Dengan mengetahui hasil analisis tersebut pemilik pasar swalayan dapat mengatur penempatan barang atau membuat strategi pemasaran untuk barang tersebut (Suwarningsih, 2008). Penerapan tersebut biasanya disebut dengan *Market Basket Analysis* karena sering digunakan dalam analisis kombinasi item pada keranjang belanjaan.

Dalam menentukan suatu *association rule*, terdapat suatu *interestingness measure* (ukuran kepercayaan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu. Umumnya ada tiga ukuran (Kusrini & Luthfi, 2009), yaitu:

- 1) *Support*: yaitu suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item/ *itemset* dari keseluruhan transaksi. Ukuran ini akan menentukan apakah suatu item/*itemset* layak untuk dicari *confidence*-nya (misal, dari seluruh transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa item A dan B dibeli bersamaan) dapat juga digunakan untuk mencari tingkat dominasi item tunggal. Secara umum dapat ditulis seperti persamaan berikut.

$$\text{Support}(A \rightarrow B) = \text{Probabilitas}(A \rightarrow B)$$

- 2) *Confidence*: yaitu suatu ukuran yang menunjukkan hubungan antar dua item secara kondisional (misal, seberapa sering item B dibeli jika orang membeli item A). Secara umum dapat ditulis seperti persamaan berikut.

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(A)}$$

- 3) *Improvement*: yaitu suatu ukuran yang menunjukkan besarnya kemungkinan dua item dapat dibeli secara bersamaan. Lebih jelasnya dapat dilihat pada persamaan berikut.

$$\text{Improvement } (A \rightarrow B) = \frac{\text{Support } (A \rightarrow B)}{\text{Support } (A) \times \text{Support } (B)}$$

2.2.9. Algoritma Apriori

Apriori adalah algoritma asli yang diusulkan oleh R. Agrawal dan R. Srikant pada tahun 1994 karena sering menganalisis *itemset* untuk aturan asosiasi *Boolean*. Algoritma Apriori merupakan algoritma untuk mengurangi ruang pencarian kombinasi item sehingga analisis dapat dilakukan lebih cepat. Selanjutnya, aturan yang dihasilkan dari algoritma apriori dapat diidentifikasi kembali untuk menentukan aturan mana yang dapat memberikan lebih banyak informasi menggunakan ukuran rasio *support*, *confidence*, dan *lift*. Kemudian, aturan asosiasi yang telah dihasilkan dapat digunakan sebagai bahan pertimbangan untuk pengambilan keputusan dalam strategi bisnis. Dua proses utama yang dilakukan dalam algoritma Apriori (Han & Kamber, 2006), adalah sebagai berikut:

1. *Join*

Dalam proses ini, setiap item digabungkan dengan hal lain sampai tidak ada lagi kombinasi yang terbentuk.

2. *Prune*

Pada proses ini, hasil dari item-item yang telah digabungkan kemudian dipangkas menggunakan minimum *support* yang telah ditentukan oleh pengguna.

BAB III

METODE PENELITIAN

3.1. Objek Penelitian

Objek penelitian ini yaitu data transaksi pengadaan dari PT XYZ. Data ini merupakan data pengadaan untuk suku cadang selama tahun 2010-2015.

3.2. Metode Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan beberapa metode antara lain:

1. Data Historis Perusahaan

Data historis merupakan data masa lalu yang dimiliki oleh sebuah perusahaan. Pada penelitian ini akan digunakan data pengadaan untuk material tertentu PT XYZ dari tahun 2010-2015 atau selama 72 bulan. Data tersebut berbentuk transaksi harian yang dilakukan oleh PT XYZ.

2. Kajian Kepustakaan

Kajian kepustakaan dilakukan untuk mendapatkan referensi terkait permasalahan serupa atau mendapatkan data sekunder untuk digunakan dalam penelitian. Selain itu kajian kepustakaan dilakukan untuk mencari tahu dasar teori terkait topik yang bersangkutan. Kajian ini dilakukan dengan mempelajari literatur-literatur seperti jurnal atau buku.

3.3. Jenis Data

Data yang digunakan dalam penelitian ini terdiri dari 1 jenis data sebagai berikut:

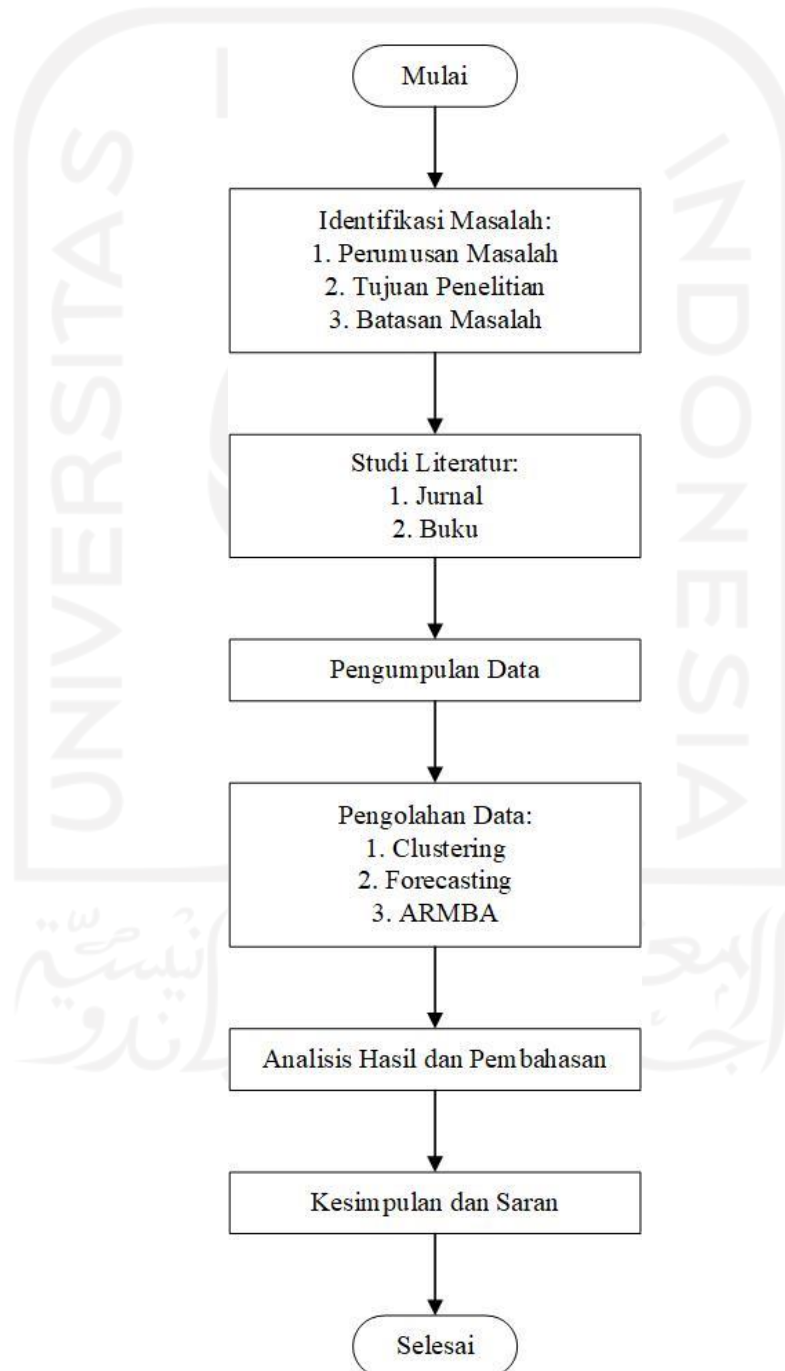
1. Data Sekunder

Data sekunder merupakan data yang tidak diambil secara langsung. Terkait penelitian akan menggunakan data sekunder berupa data historis pengadaan suku cadang PT XYZ dari tahun 2010-2015.



3.4. Alur Penelitian

Gambar 3.1. menunjukkan alur penelitian yang dilakukan beserta dengan penjelasan setiap prosesnya.



Gambar 3.1. Alur Penelitian

Penjelasan alur penelitian tersebut adalah sebagai berikut:

1. Identifikasi Masalah

Dalam identifikasi masalah akan ditentukan mengenai permasalahan yang akan dibahas atau biasa disebut dengan perumusan masalah. Dengan adanya rumusan masalah maka tujuan dari penelitian dapat ditentukan yaitu menjawab rumusan masalah yang telah ditentukan. Agar penelitian tidak melebar ke topik yang lebih luas akan ditentukan batasan masalah untuk penelitian tersebut.

2. Kajian Literatur

Kajian literatur merupakan proses mengkaji berbagai sumber literatur yang tersedia terkait dengan topik yang diangkat. Proses ini dimaksudkan untuk mencari serta memberikan landasan teori terhadap topik penelitian yang diangkat. Kajian ini terdiri dari kajian induktif dan kajian deduktif.

- a. Kajian induktif berisi penelitian-penelitian yang telah dilakukan oleh peneliti lainnya dengan topik permasalahan serupa.
- b. Kajian deduktif berisi mengenai landasan teori terkait dengan topik permasalahan maupun metode yang digunakan.

3. Pengumpulan Data

Data yang dikumpulkan merupakan data sekunder dari PT XYZ. Data ini merupakan data pengadaan suku cadang dari tahun 2010-2015 atau selama 72 bulan. Data tersebut didapatkan dari rekaman *database* PT XYZ.

4. Pengolahan Data

Pengolahan data dilakukan dengan menggunakan bantuan *Python*. Hal ini karena penggunaan *Python* memungkinkan untuk menjalankan modul *machine learning* guna mempermudah dan mempercepat proses pengolahan data. Selain itu pengolahan data dilakukan dengan menggunakan 3 metode yaitu:

- a. *Clustering*

Penggunaan *clustering* dimaksudkan untuk untuk mengelompokkan data berdasarkan kedekatan nilai. Algoritma yang digunakan yaitu *K-Means* karena algoritma tersebut mengelompokkan objek yang memiliki karakter

yang sama ke dalam satu kluster dan objek dengan karakter yang berbeda ke dalam kluster lainnya. Modul *Python* yang digunakan untuk *clustering* yaitu *scikit learn* atau *sklearn*.

b. *Forecasting*

Forecasting atau peramalan akan dilakukan dengan menggunakan metode *Holt Winter* dengan *Triple Exponential Smoothing*. Peramalan dilakukan dengan maksud untuk memprediksi tingkat kebutuhan *aggregate* dari suku cadang yang bersangkutan. Modul *Python* yang akan digunakan untuk *forecasting* yaitu *statsmodel*.

c. *Association Rules (AR)*

AR digunakan untuk mengetahui kombinasi material apa saja yang sering muncul dari transaksi yang ada. Algoritma yang digunakan yaitu algoritma apriori. Modul *Python* yang digunakan untuk AR adalah *mlxtend*.

5. Analisis Hasil dan Pembahasan

Pada proses ini akan dilakukan analisis berdasarkan dari hasil perhitungan yang sebelumnya telah dilakukan. Hal ini dilakukan untuk memahami hasil dari pengolahan data yang sebelumnya telah dilakukan.

6. Kesimpulan dan Saran

Proses ini dilakukan untuk mengambil kesimpulan dari analisis yang sebelumnya telah dilakukan serta menjawab rumusan masalah penelitian ini. Selain itu saran juga akan diberikan terkait permasalahan yang diangkat.

الجامعة الإسلامية
الاستاذ الدكتور

BAB IV

PENGOLAHAN DATA

4.1. Pengumpulan Data

Proses pengadaan yang dilakukan oleh PT XYZ meliputi banyak barang yang dibutuhkan untuk operasional. Salah satu barang tersebut adalah suku cadang untuk mesin ataupun fasilitas lain. Penelitian ini akan menggunakan data pengadaan salah satu jenis suku cadang PT XYZ. Data tersebut didapatkan dari *database* yang dimiliki oleh PT XYZ.

4.1.1. Data Historis Pengadaan

Data yang digunakan merupakan data pengadaan material dari tahun 2010-2015 dari PT XYZ. Dalam *dataset* terdapat 8 atribut yang meliputi Material No., Short Text, Purch.Doc., Item PO, PO. Date, PO Quantity, OUn, dan Crcy. Selain itu terdapat *object* atau baris sebanyak 5091 baris. Gambar 4.1. menunjukkan atribut serta tipe data dari atribut tersebut.

Material No.	int64
Short Text	object
Purch.Doc.	int64
Item PO	int64
PO. Date	datetime64[ns]
PO Quantity	int64
OUn	object
Crcy	object
dtype: object	
Dimensi: (5091, 8)	

Gambar 4.1. Atribut dan Tipe Data

Material No. merupakan kode yang diberikan kepada setiap material. Kode ini merupakan kode unik yang dimiliki oleh setiap material. *Short Text* merupakan nama dari setiap material yang dimiliki. Purch.Doc. merupakan kode transaksi pengadaan yang dilakukan. Item PO merupakan *nomor* item pada setiap transaksi. PO. Date merupakan waktu dilakukannya setiap transaksi pengadaan. PO Quantity merupakan jumlah material yang dipesan pada setiap transaksi. OUn merupakan satuan untuk setiap transaksi. Crcy merupakan mata uang yang digunakan untuk setiap transaksi. Untuk dapat digunakan dalam pengolahan data lebih lanjut, atribut serta data yang dimiliki perlu dilakukan *preprocessing* terlebih dahulu untuk menyesuaikan dengan metode yang digunakan.

4.1.2. *Preprocessing Data Historis*

Preprocessing yang dilakukan adalah pembersihan data dari atribut atau elemen yang tidak dibutuhkan serta memperkaya data dengan informasi lain yang relevan. *Preprocessing* dilakukan dengan menggunakan Microsoft Excel secara langsung. Dalam data yang digunakan, atribut atau kolom Crcy dan OUn akan dihilangkan karena tidak digunakan. Objek dari atribut Material No. dan Short Text akan diubah menjadi kode yang lebih pendek dan menjadi atribut baru dengan nama Material Code untuk penyederhanaan. Perubahan tersebut dilakukan dengan menyalin data Material No. dan Short Text lalu menghapus data yang memiliki duplikat sehingga dapat diketahui setiap material yang ada. Kode material yang baru kemudian ditambahkan pada data tersebut. Untuk menerapkan kode material yang baru dapat digunakan fungsi vlookup pada Excel. Dari pemrosesan tersebut dihasilkan 394 material *code*. Selanjutnya, PO. Date akan diuraikan menjadi atribut tanggal, bulan, dan tahun karena dibutuhkan untuk pemrosesan selanjutnya. Proses ini dilakukan dengan memisahkan objek PO. Date menjadi atribut tanggal, bulan, dan tahun. Gambar 4.2. menunjukkan mengenai atribut dan tipe data hasil *preprocessing*. Dari *preprocessing* yang telah dilakukan terdapat 7 atribut yang akan digunakan untuk proses selanjutnya.

Material Code	int64
Purch.Doc.	int64
PO. Date	datetime64[ns]
Day	int64
Month	int64
Year	int64
PO Quantity	int64
dtype:	object

Gambar 4.2. Atribut dan Tipe Data Hasil *Preprocessing*

Gambar 4.3. merupakan contoh hasil 5 data setelah dilakukan *preprocessing*. Setelah *preprocessing* terdapat atribut Material Code yang menunjukkan kode unik setiap material, *Day* yang menunjukkan tanggal transaksi, *Month* yang menunjukkan bulan transaksi, dan *Year* yang menunjukkan tahun transaksi.

	Material Code	Purch.Doc.	PO. Date	Day	Month	Year	PO Quantity
1	39	1200019111	2010-01-12	12	1	2010	8
2	45	1200019111	2010-01-12	12	1	2010	8
3	46	1200019111	2010-01-12	12	1	2010	8
4	328	1200019111	2010-01-12	12	1	2010	4
8	115	1200019208	2010-01-14	14	1	2010	2

Gambar 4.3. Contoh Data Hasil *Preprocessing*

4.2. Pengolahan Data

Pengolahan data akan dilakukan dengan 3 metode yaitu, *Clustering*, *forecasting* menggunakan *Holt Winter*, dan *Association Rule (AR)*. Dalam setiap metode akan dilakukan transformasi data. Transformasi ini dimaksudkan untuk menyesuaikan bentuk data agar dapat digunakan dalam metode yang dimaksud. Selain itu pengolahan data akan menggunakan bantuan dari *Python* dengan mengaplikasikan beberapa modul *machine learning*. Dengan *Python* pengolahan data menjadi lebih fleksibel serta cepat karena

Python dapat menjalankan banyak modul untuk melakukan analisis dan *Python* sendiri merupakan salah satu bahasa pemrograman yang populer digunakan untuk data analisis.

Import dataset hasil *preprocessing* merupakan proses yang harus dilakukan pada awal analisis. Dengan menggunakan modul *Pandas*, *import dataset* dapat dilakukan. Dikarenakan *dataset* berasal dari Excel maka *syntax* `pd.read_excel` dapat digunakan untuk *import dataset*. Gambar 4.4. merupakan *syntax* untuk melakukan *import* modul dan *dataset*. Dalam gambar tersebut *dataset* yang diimpor akan disimpan dalam variabel `df` sehingga data tersebut dapat ditampilkan atau dipanggil dengan menuliskan variabel `df`.

```
In [ ]: # Import modul
import pandas as pd

# Import dataset
df = pd.read_excel(r'E:\Kepentingan Kuliah\Tugas Akhir (TA)\Data Mining\Gasket 2010-2015_Test.xlsx',
                 sheet_name='All_Transform')
```

Gambar 4.4. *Syntax Import Modul dan Dataset*

4.2.1. *Clustering*

Proses *clustering* diawali dengan analisis multikolinearitas. Hal ini dimaksudkan untuk mencari tahu tingkat hubungan variabel yang digunakan. Hubungan yang kuat pada variabel dapat memberikan hasil akhir yang tidak akurat. Setelah proses analisis multikolinearitas dilakukan dan menunjukkan tingkat hubungan yang masih dalam batas toleransi maka dapat dilakukan proses *clustering*.

4.2.1.1. **Multikolinearitas**

Analisis multikolinearitas dilakukan dengan menggunakan modul *statmodels* pada *Python*. *Statmodels* sendiri merupakan modul *python* yang menyediakan kelas dan fungsi

untuk estimasi dari banyak model statistik, serta melakukan uji statistik, dan eksplorasi data statistik. Multikolinearitas dapat diketahui dengan menghitung nilai *Variance Inflation Factor* (VIF). Jika nilai VIF kurang dari 10 maka multikolinearitas terbilang kecil sehingga atribut dapat digunakan untuk proses *clustering*. Atribut yang akan digunakan yaitu Material Code dan PO Quantity.

Sebelum melakukan perhitungan nilai VIF, data yang dimiliki perlu dikelompokkan berdasarkan material. Hal ini karena kita ingin mengetahui hubungan material dengan jumlah kebutuhannya. Gambar 4.5. menunjukkan hasil untuk proses pengelompokan tersebut.

Material Code	PO Quantity	Freq
1	20	2
2	24	3
3	50	1
4	24	3
5	121	9

Gambar 4.5. Contoh 5 Hasil Pengelompokan

Selain itu atribut mengenai frekuensi transaksi juga akan ditambahkan. Frekuensi menunjukkan jumlah transaksi atau seberapa sering dilakukannya pengadaan untuk sebuah material. Atribut ini ditambahkan karena clustering akan dilakukan dengan menggunakan atribut Material Code, PO Quantity, dan Freq.

Setelah pengelompokan dilakukan maka dapat dilakukan perhitungan nilai VIF untuk kedua atribut tersebut. Proses ini diawali dengan melakukan impor modul *statsmodels* yang akan digunakan untuk perhitungan nilai VIF dilanjutkan untuk membuat *syntax*

perhitungan nilai VIF. Gambar 4.6. merupakan *syntax* dan hasil yang dikeluarkan untuk perhitungan VIF.

```
In [28]: vif_data = pd.DataFrame()
vif_data["Feature"] = c1.columns
vif_data["VIF"] = [variance_inflation_factor(c1.values, i)
                   for i in range(len(c1.columns))]

#print(vif_data)
vif_data
```

Out[28]:

	Feature	VIF
0	PO Quantity	3.1637
1	Freq	3.1637

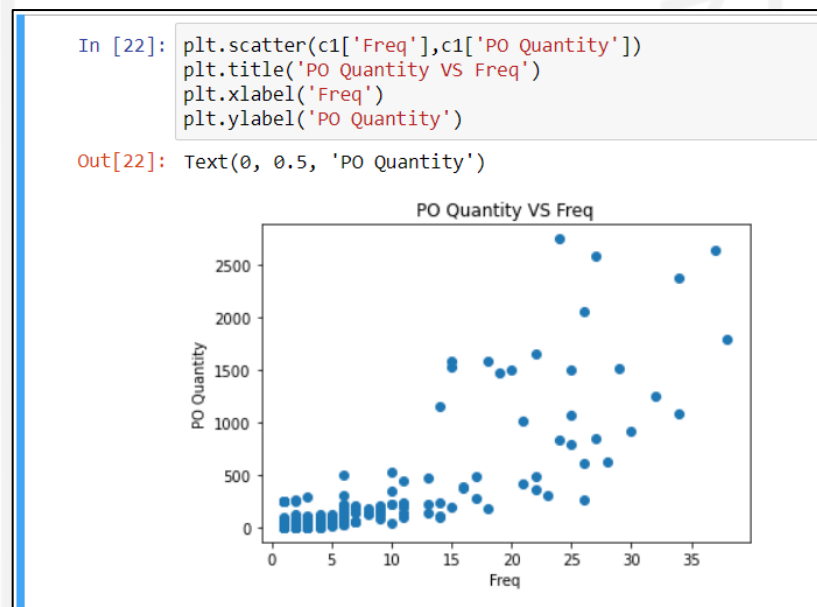
Gambar 4.6. *Syntax* dan Hasil Perhitungan VIF

Berdasarkan hasil perhitungan diketahui bahwa kedua atribut yang digunakan memiliki nilai VIF di bawah 10 atau sebesar 3.1637 untuk kedua atribut yang berarti memiliki tingkat multikolinearitas yang rendah sehingga dapat digunakan untuk proses *clustering* selanjutnya.

4.2.1.2. *K-Means*

Metode yang digunakan untuk *clustering* adalah *K-Means*. Metode ini digunakan karena jumlah kluster yang dihasilkan akan ditentukan di awal sesuai keinginan dan anggota kluster hanya tergabung dalam eksklusif 1 kluster saja. Proses *clustering* akan menggunakan modul dari *scikit-learn* atau juga sering disebut dengan *sklearn*. *Sklearn* sendiri merupakan modul *machine learning* pada *Python* yang memungkinkan untuk melakukan bermacam-macam proses klasifikasi, regresi, dan *clustering*.

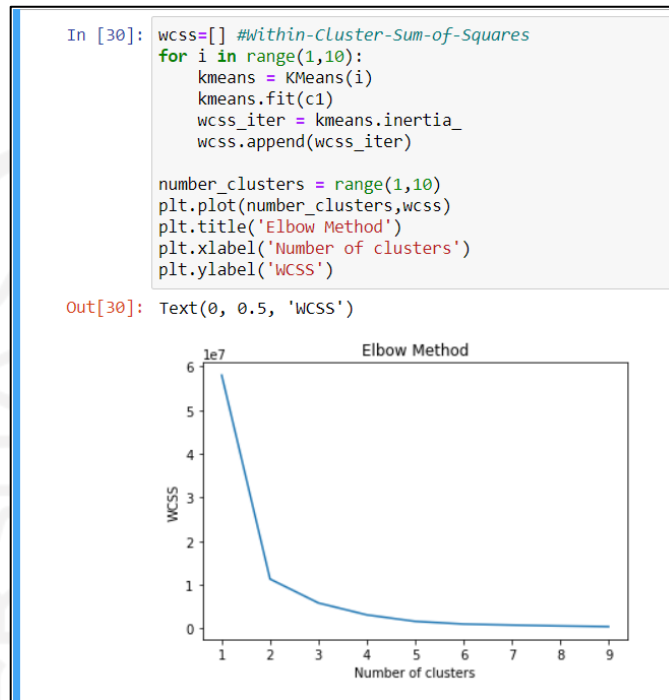
Clustering akan dilakukan berdasarkan atribut PO Quantity dan Freq karena hasil kluster diharapkan berhubungan dengan tingkat kebutuhan. Pengelompokan jumlah pengadaan berdasarkan material juga perlu untuk dilakukan sebelum proses *clustering*. Sebelum melakukan proses *clustering* dapat membuat *scatter* diagram untuk melihat distribusi dari data yang dimiliki. Pembuatan *scatter* plot dilakukan dengan menggunakan modul *matplotlib* sehingga modul tersebut harus diimpor terlebih dahulu untuk dapat digunakan. Gambar 4.7. merupakan *syntax* dan hasil *scatter* diagram yang dibuat. Pembuatan *scatter* diagram menggunakan modul *matplotlib*.



Gambar 4.7. *Syntax* dan Hasil *Scatter* Diagram

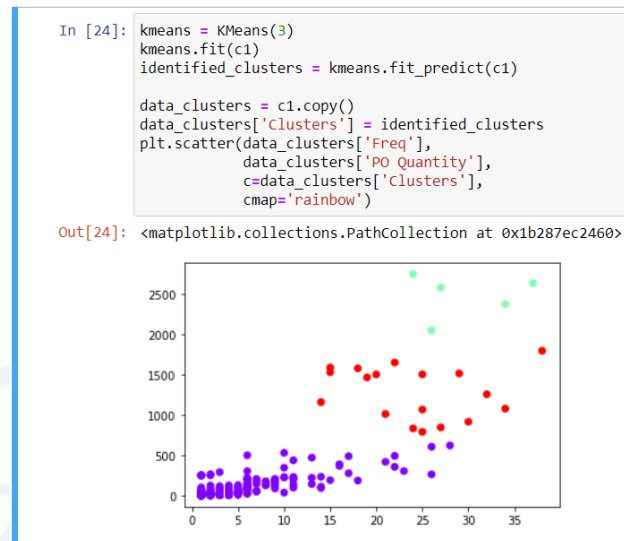
Penggunaan metode *K-Means* mengharuskan menentukan jumlah kluster pada awal atau sebelum proses pembentukan kluster. Jumlah kluster yang dihasilkan ditentukan sebanyak 3 kluster. Hal ini karena kluster yang berhubungan dengan jumlah pengadaan diharapkan dapat membagi menjadi 3 kluster. Selain itu dengan menggunakan metode *elbow*, 3 kluster merupakan jumlah kluster yang optimal untuk data yang ada. Pengambilan kesimpulan tersebut karena nilai yang ditunjukkan setelah titik ke 3 pada metode *elbow* cenderung landai, sehingga 3 kluster merupakan nilai yang optimal. Proses metode *elbow* dilakukan dengan menggunakan modul *sklearn*. Oleh karena itu modul

tersebut harus diimpor terlebih dahulu untuk dapat digunakan. Gambar 4.8. merupakan *syntax* dan hasil dari metode *elbow* untuk mengetahui jumlah kluster yang optimal.



Gambar 4.8. *Syntax* dan Hasil Metode *Elbow*

Setelah penentuan jumlah *cluster* selesai, selanjutnya akan dilakukan proses *clustering* dengan menggunakan modul *sklearn* pada *Python*. Jumlah kluster yang telah ditentukan sebelumnya juga menjadi *input* pada proses ini. Gambar 4.9. merupakan *syntax* dan hasil dari proses *clustering*.



Gambar 4.9. *Syntax* dan Hasil *Clustering*

Kluster pertama yang terbentuk memiliki 327 anggota. Tabel mengenai daftar anggota kluster yang pertama dapat dilihat pada Lampiran 1. Index kluster akan dimulai dari angka 0 karena merupakan *index* bawaan dari *Python* sehingga kluster dengan *index* 0 merupakan kluster yang pertama. Pada *scatter* diagram, kluster yang pertama ditunjukkan oleh titik berwarna ungu. Berdasarkan dari kluster yang terbentuk dapat diketahui mengenai karakteristik anggota yang tergabung. Proses ini dapat dilakukan dengan menggunakan *syntax* `dataframe.describe()` pada *Python*. *Syntax* ini akan memberikan informasi mengenai jumlah anggota (*count*), rata-rata (*mean*), standard deviasi (*std*), nilai minimum (*min*), persentil 1 (25%), persentil 2 atau median (50%), persentil 3 (75%), dan nilai maksimum (*max*). Hasil terkait karakteristik kluster yang pertama ditunjukkan oleh Gambar 4.10.

	PO Quantity	Freq	Clusters
count	327.000000	327.000000	327.0
mean	67.538226	4.067278	0.0
std	107.242247	4.416408	0.0
min	1.000000	1.000000	0.0
25%	6.000000	1.000000	0.0
50%	20.000000	3.000000	0.0
75%	79.500000	5.000000	0.0
max	621.000000	28.000000	0.0

Gambar 4.10. Karakteristik Kluster Pertama

Kluster kedua yang terbentuk memiliki 18 anggota. Tabel mengenai daftar anggota kluster yang pertama dapat dilihat pada Lampiran 2. Index kluster akan dimulai dari angka 0 karena merupakan *index* bawaan dari *Python* sehingga kluster dengan *index* 1 merupakan kluster yang kedua. Pada *scatter* diagram, kluster yang kedua ditunjukkan oleh titik berwarna merah. Berdasarkan dari kluster yang terbentuk dapat diketahui mengenai karakteristik anggota yang tergabung. Proses ini dapat dilakukan dengan menggunakan *syntax* `dataframe.describe()` pada *Python*. *Syntax* ini akan memberikan informasi mengenai jumlah anggota (*count*), rata-rata (*mean*), standard deviasi (*std*), nilai minimum (*min*), persentil 1 (25%), persentil 2 atau median (50%), persentil 3 (75%), dan nilai maksimum (*max*). Hasil terkait karakteristik kluster yang kedua ditunjukkan oleh Gambar 4.11.

	PO Quantity	Freq	Clusters
count	18.000000	18.000000	18.0
mean	1281.166667	24.055556	2.0
std	322.533035	6.795087	0.0
min	789.000000	14.000000	2.0
25%	1024.750000	19.250000	2.0
50%	1359.500000	24.500000	2.0
75%	1524.250000	28.500000	2.0
max	1794.000000	38.000000	2.0

Gambar 4.11. Karakteristik Kluster Kedua

Kluster ketiga yang terbentuk memiliki 5 anggota. Tabel mengenai daftar anggota kluster yang pertama dapat dilihat pada Lampiran 3. Index kluster akan dimulai dari angka 0 karena merupakan *index* bawaan dari *Python* sehingga kluster dengan *index* 2 merupakan kluster yang ketiga. Pada *scatter* diagram, kluster yang ketiga ditunjukkan oleh titik berwarna *cyan*. Berdasarkan dari kluster yang terbentuk dapat diketahui mengenai karakteristik anggota yang tergabung. Proses ini dapat dilakukan dengan menggunakan *syntax* `dataframe.describe()` pada *Python*. *Syntax* ini akan memberikan informasi mengenai jumlah anggota (*count*), rata-rata (*mean*), standard deviasi (*std*), nilai minimum (*min*), persentil 1 (25%), persentil 2 atau median (50%), persentil 3 (75%), dan nilai maksimum (*max*). Hasil terkait karakteristik kluster yang pertama ditunjukkan oleh Gambar 4.12.

	PO Quantity	Freq	Clusters
count	5.000000	5.00000	5.0
mean	2474.000000	29.60000	1.0
std	274.411188	5.59464	0.0
min	2047.000000	24.00000	1.0
25%	2371.000000	26.00000	1.0
50%	2576.000000	27.00000	1.0
75%	2632.000000	34.00000	1.0
max	2744.000000	37.00000	1.0

Gambar 4.12. Karakteristik Kluster Ketiga

4.2.2. *Forecasting*

Proses peramalan didahului dengan proses transformasi dari data yang dimiliki. Pada *preprocessing* data memiliki 7 atribut, namun proses peramalan hanya akan menggunakan 3 atribut saja yaitu Material Code, PO. Date, dan PO Quantity. Jika proses tersebut telah selesai maka dapat dilanjutkan dengan proses peramalan menggunakan metode Holt Winter.

4.2.2.1. **Transformasi Data**

Proses transformasi perlu dilakukan agar data yang dimiliki dapat digunakan untuk proses peramalan menggunakan metode Holt Winter. Proses peramalan akan menggunakan atribut Material Code, PO. Date, dan PO Quantity sehingga ketiga atribut ini perlu untuk dipisahkan dari atribut lainnya. Pemisahan ini dapat dilakukan dengan menggunakan *syntax* dari *Python*. Selain itu juga diperlukan pengelompokan berdasarkan bulan transaksi karena data yang diperlukan akan berbentuk agregat setiap bulannya. Pengelompokan ini juga dilakukan karena peramalan akan menggunakan

periode bulan. Gambar 4.13. merupakan *syntax* dan hasil untuk proses pemisahan atribut dan pengelompokan data yang diinginkan.

```
[10]: df_forc = df[['Material Code', 'PO. Date', 'PO Quantity']]

#Filter Cluster 2 dan 3
filter_list = ['35', '36', '45', '46', '47', '59', '61', '63', '64', '66', '67', '74', '78', '83', '85', '86', '326', '328',
              '32', '39', '69', '82', '323']
df_forc = df_forc[df_forc['Material Code'].isin(filter_list)]

#PengeLompokkan
df_forc = df_forc.drop(['Material Code'], axis=1)
df_forc = df_forc.groupby(pd.Grouper(key='PO. Date', axis=0, freq='M')).sum().dropna().sort_values('PO. Date')
df_forc = df_forc.reset_index()
df_forc = df_forc[df_forc['PO. Date'] >= '2012-01-01']
df_forc = df_forc.set_index('PO. Date')

print(df_forc.shape)
df_forc.head()

(48, 1)

[10]:
```

PO. Date	PO Quantity
2012-01-31	8
2012-02-29	8
2012-03-31	256
2012-04-30	0
2012-05-31	1174

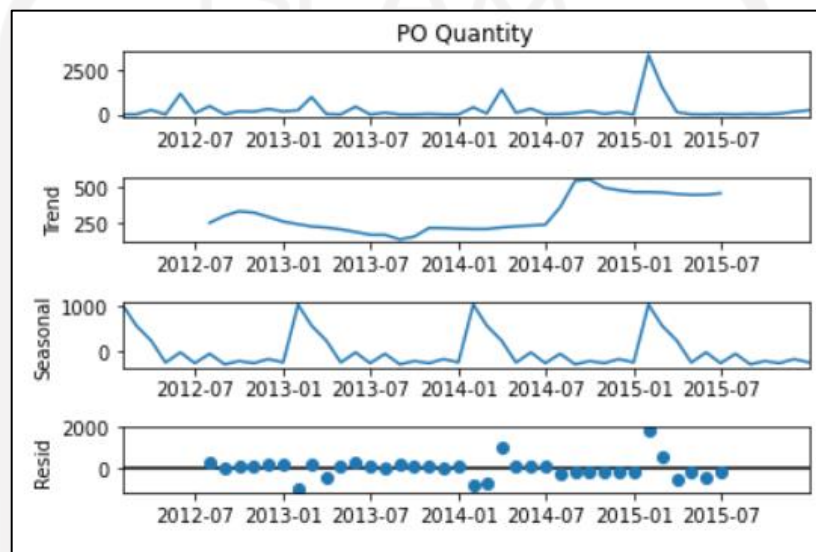
Gambar 4.13. *Syntax* dan Hasil Transformasi Data

Hasil dari pengelompokan menunjukkan data terdiri dari 48 bulan dari tahun 2012 sampai dengan 2015. Jumlah pengadaan juga telah menjadi agregat pengadaan bulanan dan tanggal pengadaan atau PO. Date menjadi *index* untuk digunakan pada pengolahan selanjutnya.

4.2.2.2. Holt Winter Seasonal

Metode forecasting kedua yang digunakan adalah Holt Winter Seasonal dengan Triple Exponential Smoothing. Dalam peramalan ini hanya berfokus pada kluster 2 dan 3 karena material yang tergabung dalam kluster tersebut bersifat slow moving dan fast moving atau dengan kata lain pengadaan dilakukan secara berkala. Kluster 1 tidak menjadi bagian dari peramalan dengan metode ini karena mayoritas material yang tergabung memiliki sifat non moving atau pengadaan tidak dilakukan secara regular. Selain itu, data tahun 2010

dan 2011 tidak ikut digunakan dalam peramalan karena distribusi data yang terbilang sangat acak sehingga akan menyulitkan dan mempengaruhi hasil akhir peramalan. Sebagai gambaran awal distribusi data dapat dilihat dengan menggunakan modul statsmodel, sehingga akan dihasilkan grafik seperti Gambar 4.14. Jika diamati dengan seksama, distribusi data tersebut menunjukkan adanya sifat seasonal atau musiman setiap awal tahun.



Gambar 4.14. Distribusi Data

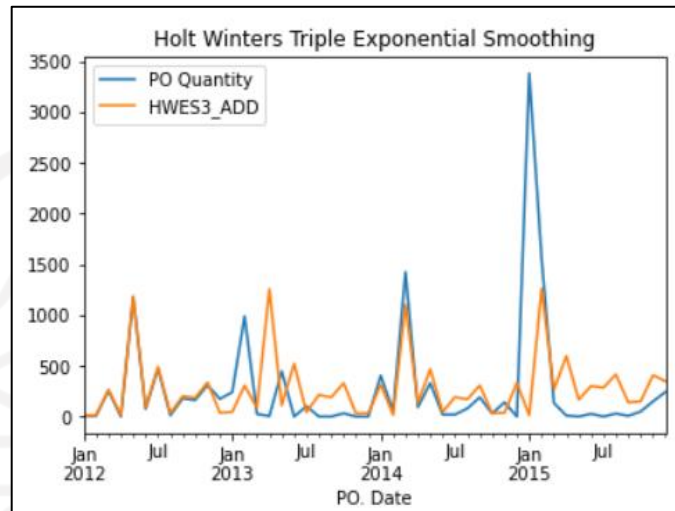
Peramalan Holt Winter akan menggunakan modul statsmodel karena modul tersebut menyediakan syntax untuk melakukan peramalan dengan metode Holt Winter. Syntax pada Gambar 4.15. merupakan proses untuk membuat model peramalan.

```
In [23]: holt['HWES3_ADD'] = ExponentialSmoothing(holt['PO Quantity'], trend='add',
                                                seasonal='add', seasonal_periods=11).fit().fittedvalues
        holt[['PO Quantity', 'HWES3_ADD']].plot(title='Holt Winters Triple Exponential Smoothing')
```

Gambar 4.15. Syntax Pembuatan Model

Perbandingan data hasil ramalan dan data aktual diperlihatkan oleh Gambar 4.16. Garis biru menunjukkan data aktual dan garis orange menunjukkan hasil peramalan.

Dengan adanya grafik ini dapat diketahui perbandingan dan variansi dari data aktual dan hasil peramalan.



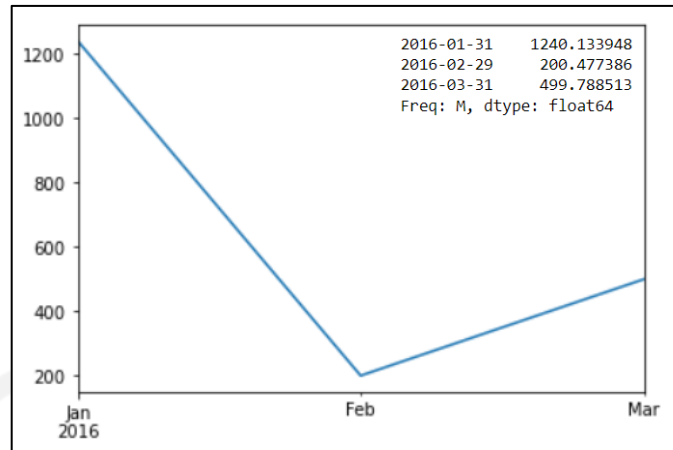
Gambar 4.16. Perbandingan Data Aktual dan Peramalan

Dalam peramalan tersebut terdapat variasi antar nilai aktual dan peramalan atau yang sering disebut dengan error peramalan. Gambar 4.17. menunjukkan tingkat error dari hasil peramalan dengan menggunakan mean absolute error.

```
In [131]: mean_absolute_error(holt['PO Quantity'],holt['HWES3_ADD'])
Out[131]: 244.8056731614952
```

Gambar 4.17. Tingkat Error Peramalan

Peramalan untuk 3 bulan mendatang dari bulan Januari sampai dengan Maret 2016 ditunjukkan oleh Gambar 4.18. Hasil peramalan untuk bulan Januari sebesar 1241, Februari sebesar 201, dan Maret sebesar 500.



Gambar 4.18. Hasil Peramalan 3 Bulan

4.2.2.3. Prediksi Material

Peramalan terkait jumlah kebutuhan telah dilakukan dengan metode Holt Winter. Hasil dari peramalan belum dapat menggambarkan material yang diperlukan. Oleh karena itu, analisis lebih lanjut dapat dilakukan untuk memprediksi material apa saja yang diperlukan setiap bulannya. Analisis ini dapat dilakukan dengan menggunakan peluang dari munculnya material setiap bulan. Penggunaan *Python* memungkinkan untuk mencari peluang tersebut. Dengan menggunakan *syntax* seperti Gambar 4.19., proses analisis peluang dapat dilakukan. Proses yang dilakukan *syntax* ini adalah dengan mencari tahu kemunculan dari sebuah setiap bulannya selama tahun 2010 sampai dengan 2015. Maka setelah itu dapat diketahui peluang munculnya material untuk tahun berikutnya.

```
In [363]: def most (month,limit):
x = df[['Material Code','Month','Year','PO Quantity']]

x = x[x['Month']==month]

x = (x.groupby(['Material Code','Year'])['PO Quantity']
     .sum().unstack().reset_index().fillna(0)
     .set_index('Material Code'))

def hot_encode(x):
    if(x<= 0):
        return 0
    if(x>= 1):
        return 1

all_encoded = x.applymap(hot_encode)
x = all_encoded

y = x.copy()
y["Sum"] = x.sum(axis=1)

return y[y['Sum']>=limit]
```

Gambar 4.19. *Syntax* Perhitungan Peluang

Hasil untuk prediksi material pada bulan Januari ditunjukkan oleh Tabel 4.1. Hasil ini didapat dengan menjalankan fungsi `most(1,3)` yang telah dibuat pada *syntax* sebelumnya. Peluang untuk keluarnya material pada bulan Januari ditunjukkan oleh kolom `Sum`. Kolom ini menunjukkan jumlah sebuah material muncul pada setiap tahunnya. Minimal nilai yang ditentukan untuk pengambilan nilai `Sum` ini adalah 3 atau sebesar 50%. Sehingga, material yang berkemungkinan muncul atau dibutuhkan dengan tingkat peluang 50% pada bulan Januari sebanyak 10 material.

Tabel 4.1. Hasil Perhitungan Peluang Bulan Januari

Material Code	2010	2011	2012	2013	2014	2015	Sum
39	1	0	0	1	1	0	3
45	1	0	0	1	1	0	3
46	1	0	0	1	1	0	3
47	1	1	0	1	1	0	4
67	0	1	0	0	1	1	3
71	0	1	0	0	1	1	3
74	0	1	0	0	1	1	3
82	1	1	0	1	1	1	5
83	1	1	1	1	1	1	6
152	0	1	0	1	0	1	3

Hasil untuk prediksi material pada bulan Februari ditunjukkan oleh Tabel 4.2. Hasil ini didapat dengan menjalankan fungsi *most(2,3)* yang telah dibuat pada *syntax* sebelumnya. Peluang untuk keluarnya material pada bulan Februari ditunjukkan oleh kolom Sum. Kolom ini menunjukkan jumlah sebuah material muncul pada setiap tahunnya. Minimal nilai yang ditentukan untuk pengambilan nilai Sum ini adalah 3 atau sebesar 50%. Sehingga, material yang berkemungkinan muncul atau dibutuhkan dengan tingkat peluang 50% pada bulan Februari sebanyak 22 material.

Tabel 4.2. Hasil Perhitungan Peluang Bulan Februari

Material Code	2010	2011	2012	2013	2014	2015	Sum
47	0	1	0	1	1	1	4
49	1	1	0	1	0	1	4
50	1	1	0	1	0	0	3
52	1	1	0	1	0	0	3
64	1	1	0	1	0	0	3
67	1	1	0	1	0	0	3
69	1	1	0	1	0	0	3
74	1	1	0	1	0	1	4
78	0	1	0	1	1	0	3
79	1	1	0	0	0	1	3
82	1	1	0	1	1	0	4
83	1	1	1	1	0	0	4
85	1	1	0	1	1	0	4
86	1	1	0	1	0	1	4
112	1	1	0	1	0	0	3
194	0	1	0	0	1	1	3
197	0	1	0	1	0	1	3
199	0	1	0	1	0	1	3
212	0	1	0	0	1	1	3
323	1	1	0	0	0	1	3
328	1	1	0	1	0	1	4
329	0	1	0	1	0	1	3

Hasil untuk prediksi material pada bulan Maret ditunjukkan oleh Tabel 4.3. Hasil ini didapat dengan menjalankan fungsi *most(3,3)* yang telah dibuat pada *syntax* sebelumnya. Peluang untuk keluarnya material pada bulan Maret ditunjukkan oleh kolom Sum. Kolom ini menunjukkan jumlah sebuah material muncul pada setiap tahunnya. Minimal nilai

yang ditentukan untuk pengambilan nilai Sum ini adalah 3 atau sebesar 50%. Sehingga, material yang berkemungkinan muncul atau dibutuhkan dengan tingkat peluang 50% pada bulan Maret sebanyak 20 material.

Tabel 4.3. Hasil Perhitungan Peluang Bulan Maret

Material Code	2010	2011	2012	2013	2014	2015	Sum
32	1	0	1	0	1	0	3
36	1	1	0	1	1	0	4
39	1	1	1	1	1	1	6
45	1	0	0	1	1	0	3
46	1	0	0	1	1	1	4
47	1	0	1	1	1	1	5
51	1	0	1	0	1	1	4
52	1	1	0	1	1	1	5
69	1	0	0	0	1	1	3
71	1	0	0	0	1	1	3
78	1	1	0	0	1	0	3
82	1	1	0	0	1	1	4
83	1	0	1	0	1	1	4
86	1	1	0	0	1	1	4
114	1	1	1	0	1	0	4
196	0	0	0	1	1	1	3
197	0	1	0	1	1	1	4
326	1	1	0	1	0	0	3
328	1	1	1	0	1	1	5
334	1	1	0	0	1	0	3

4.2.3. Association Rule (AR)

Proses AR didahului dengan proses transformasi dari data yang dimiliki. Bentuk data tersebut akan diubah untuk menggambarkan dalam satu transaksi terdapat item apa saja. Jika proses tersebut telah selesai maka dapat dilanjutkan dengan proses AR.

4.2.3.1. Transformasi Data

Sebelum melakukan analisis AR data yang dimiliki perlu dilakukan penyesuaian atau transformasi data. Hal ini dimaksudkan agar bentuk data dapat digunakan untuk proses selanjutnya. Dengan menggunakan *syntax* seperti Gambar 4.20., data yang dimiliki dapat sesuai untuk digunakan pada proses selanjutnya. *Syntax* ini akan mengelompokkan jenis material berdasarkan Purch.Doc. atau kode transaksi sehingga dapat diketahui material apa saja yang terdapat dalam sebuah transaksi.

```
In [ ]: # Transformasi Data
all_data = (df.groupby(['Purch.Doc.', 'Material Code'])['PO Quantity']
            .sum().unstack().reset_index().fillna(0)
            .set_index('Purch.Doc.'))

def hot_encode(x):
    if(x<= 0):
        return 0
    if(x>= 1):
        return 1

all_encoded = all_data.applymap(hot_encode)
all_data = all_encoded
```

Gambar 4.20. *Syntax* Transformasi Data

Gambar 4.21. merupakan hasil transformasi data yang dilakukan. Setelah dilakukan dapat diketahui material apa saja yang terdapat dalam sebuah transaksi. Selain juga diketahui bahwa jumlah Purch.Doc. atau transaksi sebanyak 269 transaksi.

Material Code	1	10	100	102	103	104	106	108	109	11	...	9	90	91	92	94	95	96	97	98	99	
Purch.Doc.																						
1200017706	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1200019111	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1200019208	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1200019524	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1200019765	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
...
1200153091	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1200153731	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1200154661	0	0	0	0	0	0	0	0	0	0	...	1	0	0	0	0	0	0	0	0	0	0
1200154838	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1200155333	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0

269 rows × 350 columns

Gambar 4.21. Hasil Transformasi

4.2.3.2. Algoritma Apriori

Proses AR dilakukan dengan menggunakan algoritma apriori. Proses tersebut juga akan menggunakan modul *mlxtend* pada *Python* karena modul tersebut menyediakan algoritma apriori. Parameter yang digunakan untuk mencari aturan AR adalah nilai *support* dan *confidence* dari kombinasi material. Ditentukan nilai minimal *support* yang digunakan yaitu sebesar 10% dan minimal *confidence* sebesar 50%. Gambar 4.22. merupakan *syntax* untuk proses AR semua transaksi.

```
In [15]: # Import Modul
from mlxtend.frequent_patterns import apriori, association_rules

# Pembuatan Model
frq_items = apriori(all_data, min_support = 0.1, use_colnames = True)

# Proses ARMB
rules = association_rules(frq_items, metric = "confidence", min_threshold = 0.5)
rules = rules.sort_values(['confidence', 'lift'], ascending = [False, False])
print(rules.shape)
rules
```

Gambar 4.22. Syntax Pembuatan AR

Hasil AR atau kombinasi material untuk semua transaksi ditunjukkan oleh Tabel 4.4. Secara keseluruhan terbentuk sebanyak 24 aturan AR dengan minimal *support* sebesar 10% dan minimal *confidence* sebesar 50%.

Tabel 4.4. Hasil Aturan AR

Antecedents	Consequents	Support	Confidence	Lift
{'78'}	{'82'}	0.1301	0.7778	3.0322
{'85'}	{'82'}	0.1115	0.7692	2.9989
{'85'}	{'83'}	0.1115	0.7692	2.9560
{'51'}	{'328'}	0.1004	0.7297	3.5690
{'86'}	{'83'}	0.1115	0.6977	2.6811
{'66'}	{'82'}	0.1004	0.6923	2.6990
{'45'}	{'46'}	0.1264	0.6415	3.0815
{'49'}	{'83'}	0.1041	0.6364	2.4455
{'46'}	{'47'}	0.1301	0.6250	2.5473
{'46'}	{'45'}	0.1264	0.6071	3.0815
{'45'}	{'47'}	0.1190	0.6038	2.4608
{'46'}	{'39'}	0.1152	0.5536	2.5239
{'82'}	{'83'}	0.1413	0.5507	2.1164
{'83'}	{'82'}	0.1413	0.5429	2.1164
{'69'}	{'46'}	0.1004	0.5400	2.5939
{'69'}	{'82'}	0.1004	0.5400	2.1052
{'47'}	{'46'}	0.1301	0.5303	2.5473
{'45'}	{'82'}	0.1041	0.5283	2.0596
{'328'}	{'47'}	0.1078	0.5273	2.1490
{'39'}	{'46'}	0.1152	0.5254	2.5239
{'45'}	{'39'}	0.1004	0.5094	2.3227
{'39'}	{'47'}	0.1115	0.5085	2.0724
{'82'}	{'78'}	0.1301	0.5072	3.0322
{'47'}	{'83'}	0.1227	0.5000	1.9214

BAB V

PEMBAHASAN

5.1. *Clustering*

Clustering merupakan metode pengelompokan yang termasuk dalam *unsupervised learning*. Hal ini berarti sistem dapat melakukan identifikasi pola dengan sendirinya. Metode ini digunakan karena keterbatasan data pendukung atau parameter-parameter untuk mengelompokkan data. Oleh karena itu, metode ini diharapkan dapat mengelompokkan data yang ada berdasarkan kedekatan nilai. Terkait data yang digunakan merupakan data pengadaan material pada tahun 2010-2015 dan atribut yang digunakan yaitu Material Code dan PO Quantity karena hasil *clustering* diharapkan dapat menggambarkan kelompok material berdasarkan jumlah pengadaan yang selama ini dilakukan. Jumlah kluster juga ditentukan di awal atau sebanyak 3 kluster karena *clustering* menggunakan metode *K-Means* sehingga mengharuskan penentuan jumlah kluster di awal perhitungan. Penentuan jumlah kluster tersebut juga didukung hasil dari metode *elbow* yang menunjukkan bahwa 3 kluster merupakan jumlah yang optimal. Pengambilan kesimpulan jumlah kluster tersebut didasarkan dari nilai *Within Cluster Sum of Squares* (WCSS) yang menunjukkan nilai WCSS pada jumlah kluster 4 tidak memberikan perbedaan nilai yang signifikan dibandingkan dengan jumlah kluster 3. Oleh karena itu 3 kluster merupakan jumlah yang optimal untuk data tersebut.

Hasil kluster yang pertama memiliki anggota 327 suku cadang. Dalam kluster ini terdiri dari 2 atribut yaitu PO Quantity dan Freq. PO Quantity menggambarkan jumlah material yang dipesan dan Freq menggambarkan frekuensi transaksi yang dilakukan selama 72 bulan. Dengan adanya kedua atribut tersebut maka dapat diketahui apakah sebuah material dilakukan pengadaan secara berulang atau hanya sekali saja, dengan kata

lain material yang bersifat *fast moving*, *slow moving*, dan *non moving* dapat diketahui dengan pasti. Berdasarkan PO Quantity, karakteristik dari kluster ini memiliki nilai rata-rata sebesar 67.54, nilai standar deviasi sebesar 107.24, nilai minimum sebesar 1, persentil 1 sebesar 6, persentil 2 atau median sebesar 20, persentil 3 sebesar 79.5, dan nilai maksimum sebesar 621. Oleh karena itu kluster ini dapat dikatakan kluster *low demand* karena memiliki rata-rata terkecil dibandingkan dengan kluster lainnya. Selain itu nilai minimum dan maksimum menunjukkan bahwa selama 72 bulan kebutuhan terkecil sebesar 1 suku cadang dan kebutuhan terbesar sebesar 621 suku cadang. Berdasarkan Freq, karakteristik dari kluster ini memiliki nilai rata-rata sebesar 4.1, nilai standar deviasi sebesar 4.4, nilai minimum sebesar 1, persentil 1 sebesar 1, persentil 2 atau median sebesar 3, persentil 3 sebesar 5, dan nilai maksimum sebesar 28. Oleh karena itu kluster ini dapat dikatakan kluster *non moving* karena memiliki rata-rata terkecil dibandingkan dengan kluster lainnya dan mayoritas material yang tergabung dalam kluster ini tidak melakukan pengadaan secara reguler jika dilihat dari nilai rata-rata frekuensi pengadaannya. Dengan diketahui nilai tersebut proses serta strategi pengadaan dapat direncanakan dengan lebih baik karena memiliki gambaran terkait jumlah kebutuhan selama 72 bulan sebelumnya.

Hasil kluster yang kedua memiliki anggota 18 material. Dalam kluster ini terdiri dari 2 atribut yaitu PO Quantity dan Freq. PO Quantity menggambarkan jumlah material yang dipesan dan Freq menggambarkan frekuensi transaksi yang dilakukan selama 72 bulan. Dengan adanya kedua atribut tersebut maka dapat diketahui apakah sebuah material dilakukan pengadaan secara berulang atau hanya sekali saja, dengan kata lain material yang bersifat *fast moving*, *slow moving*, dan *non moving* dapat diketahui dengan pasti. Berdasarkan PO Quantity, karakteristik dari kluster ini memiliki nilai rata-rata sebesar 1281.17, nilai standar deviasi sebesar 322.5, nilai minimum sebesar 789, persentil 1 sebesar 1024.75, persentil 2 atau median sebesar 1359.5, persentil 3 sebesar 1524.25, dan nilai maksimum sebesar 1794. Oleh karena itu kluster ini dapat dikatakan kluster *middle demand* karena memiliki rata-rata di pertengahan dibandingkan dengan kluster lainnya. Selain itu nilai minimum dan maksimum menunjukkan bahwa selama 72 bulan kebutuhan terkecil sebesar 789 suku cadang dan kebutuhan terbesar sebesar 1794 suku cadang. Berdasarkan Freq, karakteristik dari kluster ini memiliki nilai rata-rata sebesar 24.1, nilai

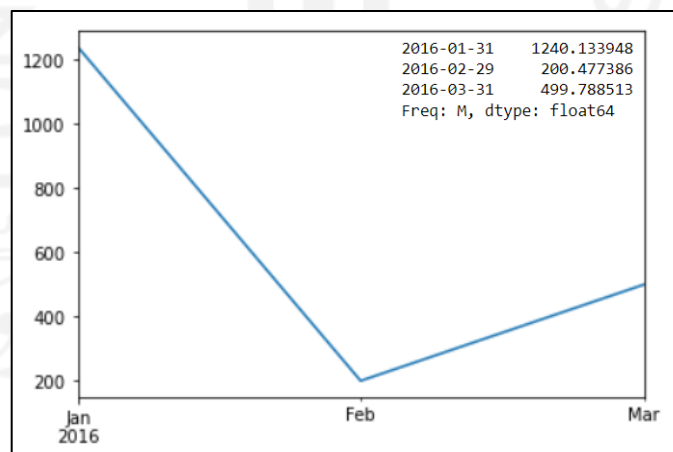
standar deviasi sebesar 6.8, nilai minimum sebesar 14, persentil 1 sebesar 19.25, persentil 2 atau median sebesar 24.5, persentil 3 sebesar 28.5, dan nilai maksimum sebesar 38. Oleh karena itu kluster ini dapat dikatakan kluster *slow moving* karena memiliki rata-rata dipertengahan dibandingkan dengan kluster lainnya dan mayoritas material yang tergabung dalam kluster ini melakukan pengadaan secara regular jika dilihat dari nilai rata-rata frekuensi pengadaannya. Dengan diketahui nilai tersebut proses serta strategi pengadaan dapat direncanakan dengan lebih baik karena memiliki gambaran terkait jumlah kebutuhan selama 72 bulan sebelumnya.

Hasil kluster yang ketiga memiliki anggota 5 material. Dalam kluster ini terdiri dari 2 atribut yaitu PO Quantity dan Freq. PO Quantity menggambarkan jumlah material yang dipesan dan Freq menggambarkan frekuensi transaksi yang dilakukan selama 72 bulan. Dengan adanya kedua atribut tersebut maka dapat diketahui apakah sebuah material dilakukan pengadaan secara berulang atau hanya sekali saja, dengan kata lain material yang bersifat *fast moving*, *slow moving*, dan *non moving* dapat diketahui dengan pasti. Berdasarkan PO Quantity, karakteristik dari kluster ini memiliki nilai rata-rata sebesar 2474, nilai standar deviasi sebesar 274.4, nilai minimum sebesar 2047, persentil 1 sebesar 2371, persentil 2 atau median sebesar 2576, persentil 3 sebesar 2632, dan nilai maksimum sebesar 2744. Oleh karena itu kluster ini dapat dikatakan kluster *high demand* karena memiliki rata-rata terkecil dibandingkan dengan kluster lainnya. Selain itu nilai minimum dan maksimum menunjukkan bahwa selama 72 bulan kebutuhan terkecil sebesar 2047 suku cadang dan kebutuhan terbesar sebesar 2744 suku cadang. Berdasarkan Freq, karakteristik dari kluster ini memiliki nilai rata-rata sebesar 29.6, nilai standar deviasi sebesar 5.6, nilai minimum sebesar 24, persentil 1 sebesar 26, persentil 2 atau median sebesar 27, persentil 3 sebesar 34, dan nilai maksimum sebesar 37. Oleh karena itu kluster ini dapat dikatakan kluster *fast moving* karena memiliki rata-rata terbesar dibandingkan dengan kluster lainnya dan mayoritas material yang tergabung dalam kluster ini melakukan pengadaan secara regular jika dilihat dari nilai rata-rata frekuensi pengadaannya. Dengan diketahui nilai tersebut proses serta strategi pengadaan dapat direncanakan dengan lebih baik karena memiliki gambaran terkait jumlah kebutuhan selama 72 bulan sebelumnya.

Hasil ini dapat digunakan sebagai bahan pertimbangan tambahan dalam pengambilan keputusan. Sebagai contoh hasil ini dapat menjadi pembagian prioritas dalam proses pengadaan serta dapat menjadi faktor pemberat jika digabungkan dengan metode lainnya seperti *forecasting*. Selain itu dengan diketahuinya nilai minimum dan maksimum maka penyesuaian terhadap jumlah pengadaan dapat dilakukan. Dengan mempertimbangkan hal tersebut biaya pemesanan dan penyimpanan dapat diefisienkan.

5.2. *Forecasting*

Tujuan dari peramalan adalah memprediksi dari nilai masa depan. Dengan adanya peramalan memungkinkan untuk seseorang mendapat gambaran masa depan dan bersiap atau menentukan strategi terkait hal tersebut. Dalam penelitian ini dilakukan proses peramalan untuk mengetahui prediksi dari jumlah kebutuhan pengadaan untuk material yang ada. Periode yang digunakan juga Metode yang digunakan yaitu *Holt Winter Seasonal*.



Gambar 5.1. Hasil Peramalan

Hasil dari peramalan ditunjukkan oleh Gambar 5.1. Pada bulan Januari prediksi kebutuhan adalah sebesar 1241, bulan Februari sebesar 201, dan bulan Maret sebesar 500.

Nilai *error* dari hasil perhitungan MAE juga menunjukkan nilai yang cukup besar atau sebesar 245. *Error* yang cukup besar ini mungkin diakibatkan oleh distribusi data yang tidak beraturan. Hal ini mungkin saja diakibatkan oleh agregasi terhadap jumlah material yang ada sehingga membuat data menjadi bias. Walaupun hasil peramalan menunjukkan *error* yang cukup besar, namun data peramalan ini menunjukkan suatu kesamaan yaitu adanya *peak* atau lonjakan pengadaan yang signifikan terjadi pada awal bulan, sehingga hal ini dapat menjadi perhatian atau peringatan kepada *user* untuk membuat strategi pengadaan pada awal bulan.

Di lain sisi, hasil peramalan ini berbentuk agregat atau akumulasi dari semua material yang ada. Oleh karena itu, analisis untuk mengetahui material apa saja yang berkemungkinan untuk muncul atau dibutuhkan pada 3 bulan peramalan perlu untuk dilakukan. Salah satu cara yang dapat digunakan untuk mengetahui material yang mungkin untuk muncul adalah dengan menggunakan peluang kemunculan. Proses perhitungan peluang yaitu dengan memeriksa kemunculan setiap material pada bulan yang diinginkan. Bulan atau periode pengecekan sama dengan periode peramalan atau pada bulan Januari, Februari, dan Maret. Selain itu juga ditentukan minimal nilai peluang sebesar 50%.

Pada bulan Januari material yang memiliki peluang untuk muncul sebesar 50% atau lebih yaitu sebanyak 10 material. Material yang memiliki peluang 50% adalah material dengan kode 39, 45, 46, 67, 71, 74, dan 152. Dengan kata lain material tersebut muncul pada setengah periode data. Material yang memiliki peluang 66.66% adalah material dengan kode 47. Dengan kata lain material tersebut muncul pada 4 periode data. Material yang memiliki peluang 83.33% adalah material dengan kode 82. Dengan kata lain material tersebut muncul pada 5 periode data. Material yang memiliki peluang 100% adalah material dengan kode 83. Dengan kata lain material tersebut muncul pada semua periode data.

Pada bulan Februari material yang memiliki peluang untuk muncul sebesar 50% atau lebih yaitu sebanyak 22 material. Material yang memiliki peluang 50% adalah material dengan kode 50, 52, 64, 67, 69, 78, 79, 112, 194, 197, 199, 212, 323, dan 329. Dengan kata lain material tersebut muncul pada setengah periode data. Material yang memiliki peluang 66.66% adalah material dengan kode 47, 49, 74, 82, 83, 85, 86, dan 328.

Pada bulan Maret material yang memiliki peluang untuk muncul sebesar 50% atau lebih yaitu sebanyak 20 material. Material yang memiliki peluang 50% adalah material dengan kode 32, 45, 69, 71, 78, 196, 326, dan 334. Dengan kata lain material tersebut muncul pada setengah periode data. Material yang memiliki peluang 66.66% adalah material dengan kode 36, 46, 51, 82, 83, 86, 114, dan 197. Dengan kata lain material tersebut muncul pada 4 periode data. Material yang memiliki peluang 83.33% adalah material dengan kode 47, 52, dan 328. Dengan kata lain material tersebut muncul pada 5 periode data. Material yang memiliki peluang 100% adalah material dengan kode 39. Dengan kata lain material tersebut muncul pada semua periode data.

Hasil ini akan dapat memberikan gambaran mengenai kebutuhan material untuk 3 bulan ke depan. Selain itu hasil *clustering* juga dapat digunakan untuk memberi pertimbangan tambahan pada material yang memiliki peluang muncul 50%. *Clustering* akan memberikan informasi terkait kluster dan karakteristik di mana material tersebut tergabung.

5.3. *Association Rule (AR)*

Analisis AR dimaksudkan untuk mencari tahu kombinasi material yang sering muncul dari setiap transaksi yang ada. Pada AR sendiri terdapat beberapa parameter yang dapat digunakan untuk mengetahui tingkat kombinasi yaitu *support*, *confidence*, dan *lift*. Adanya parameter ini dapat digunakan oleh *user* untuk mendukung pengambilan keputusan yang mereka lakukan.

Analisis AR dilakukan tidak berdasarkan periode, namun menggunakan sebanyak 269 dokumen transaksi yang ada. Berdasarkan data tersebut, terbentuk 24 aturan dengan nilai minimal *support* sebesar 10% dan nilai minimal *confidence* sebesar 50%. Nilai minimal ini ditentukan agar aturan yang dihasilkan memiliki kecenderungan terjadi menjadi lebih besar. Jika nilai minimal *confidence* ditentukan lebih kecil dari 50% maka sebagian aturan yang muncul akan cenderung tidak terjadi. Sebagai contoh aturan yang memiliki nilai *confidence* paling tinggi adalah kombinasi material 78 dan 82. Aturan ini memiliki nilai *confidence* sebesar 77.78% yang berarti jika pembelian material 78 dilakukan maka 77.78% material 82 juga ikut serta dalam transaksi tersebut. Selain itu aturan ini memiliki nilai *support* 13.01% dan nilai lift 3.0322 yang berarti transaksi untuk kedua material tersebut muncul pada 13.01% dari keseluruhan transaksi atau muncul pada 34 transaksi dan meningkatkan peluang pembelian material 82 sebesar 3.0322 kali jika material 78 telah dibeli. Tabel 5.1. merupakan daftar lengkap mengenai 24 aturan yang terbentuk beserta nilai *support*, *confidence*, dan lift.

Tabel 5.1. Hasil Aturan AR

Antecedents	Consequents	Support	Confidence	Lift
{'78'}	{'82'}	0.1301	0.7778	3.0322
{'85'}	{'82'}	0.1115	0.7692	2.9989
{'85'}	{'83'}	0.1115	0.7692	2.9560
{'51'}	{'328'}	0.1004	0.7297	3.5690
{'86'}	{'83'}	0.1115	0.6977	2.6811
{'66'}	{'82'}	0.1004	0.6923	2.6990
{'45'}	{'46'}	0.1264	0.6415	3.0815
{'49'}	{'83'}	0.1041	0.6364	2.4455
{'46'}	{'47'}	0.1301	0.6250	2.5473
{'46'}	{'45'}	0.1264	0.6071	3.0815
{'45'}	{'47'}	0.1190	0.6038	2.4608
{'46'}	{'39'}	0.1152	0.5536	2.5239
{'82'}	{'83'}	0.1413	0.5507	2.1164
{'83'}	{'82'}	0.1413	0.5429	2.1164
{'69'}	{'46'}	0.1004	0.5400	2.5939
{'69'}	{'82'}	0.1004	0.5400	2.1052
{'47'}	{'46'}	0.1301	0.5303	2.5473
{'45'}	{'82'}	0.1041	0.5283	2.0596
{'328'}	{'47'}	0.1078	0.5273	2.1490

{'39'}	{'46'}	0.1152	0.5254	2.5239
{'45'}	{'39'}	0.1004	0.5094	2.3227
{'39'}	{'47'}	0.1115	0.5085	2.0724
{'82'}	{'78'}	0.1301	0.5072	3.0322
{'47'}	{'83'}	0.1227	0.5000	1.9214

5.4. Integrasi

Hasil dari ketiga metode yang digunakan dapat diintegrasikan untuk mencapai suatu tujuan yang diinginkan. Dalam penelitian ini salah satu tujuan yang ingin dicapai adalah mendapat gambaran terkait kebutuhan material masa depan. Hal ini dapat dimulai dengan melihat pada hasil peramalan. Adanya peramalan akan menggambarkan mengenai jumlah material yang dibutuhkan secara agregat. Gambar 5.2. merupakan hasil peramalan kebutuhan material.

2016-01-31	1240.133948
2016-02-29	200.477386
2016-03-31	499.788513
Freq: M, dtype: float64	

Gambar 5.2. Hasil Peramalan

Namun hasil tersebut belum dapat memberikan gambaran mengenai material apa saja yang dibutuhkan pada bulan terkait. Oleh karena itu, penggunaan peluang dapat digunakan untuk melengkapi kekurangan tersebut. Nilai minimum peluang yang ditentukan yaitu sebesar 50% karena nilai tersebut akan dapat memberikan gambaran kecenderungan munculnya suatu material. Tabel 5.2., Tabel 5.3., dan Tabel 5.4. di bawah ini merupakan daftar material dan peluangnya untuk muncul pada bulan Januari, Februari, dan Maret.

Tabel 5.2. Hasil Peluang Bulan Januari

Januari	
Material Code	Probabilitas
39	50.0%
45	50.0%
46	50.0%
47	66.7%
67	50.0%
71	50.0%
74	50.0%
82	83.3%
83	100.0%
152	50.0%

Tabel 5.3. Hasil Peluang Bulan Februari

Februari	
Material Code	Probabilitas
47	66.7%
49	66.7%
50	50.0%
52	50.0%
64	50.0%
67	50.0%
69	50.0%
74	66.7%
78	50.0%
79	50.0%
82	66.7%
83	66.7%
85	66.7%
86	66.7%
112	50.0%
194	50.0%
197	50.0%
199	50.0%
212	50.0%
323	50.0%
328	66.7%
329	50.0%

Tabel 5.4. Hasil Peluang Bulan Maret

Maret	
Material Code	Probabilitas
32	50.0%
36	66.7%
39	100.0%
45	50.0%
46	66.7%
47	83.3%
51	66.7%
52	83.3%
69	50.0%
71	50.0%
78	50.0%
82	66.7%
83	66.7%
86	66.7%
114	66.7%
196	50.0%
197	66.7%
326	50.0%
328	83.3%
334	50.0%

Jika kita mengamati hasil tersebut dengan seksama akan ditemukan nilai peluang yang sama dengan 50%. Nilai ini belum dapat memberikan gambaran mengenai kecenderungan kejadian sebuah material karena nilai ini berada pada pertengahan. Untuk dapat menutupi kekurangan ini dapat memasukkan hasil dari pengolahan *clustering*. *Clustering* akan dapat menunjukkan karakteristik dari sebuah material. Sebagai contoh jika sebuah material memiliki peluang 50% untuk muncul dan bergabung dalam kluster ketiga maka material tersebut memiliki faktor pemberat karena kluster ketiga merupakan kluster dengan rata-rata tertinggi atau dapat dikatakan *high demand* dan *fast moving*. Kluster kedua juga merupakan faktor pemberat. Namun, untuk kluster pertama menjadi faktor yang melemahkan karena karakter dari kluster ini yaitu *low demand* dan *non moving*. Oleh karena itu, material yang bergabung dalam kluster kedua (*index 1*) dan ketiga (*index 2*) direkomendasikan untuk dilakukan pengadaan dan suku cadang yang bergabung dalam kluster pertama (*index 0*) tidak direkomendasikan untuk dilakukan pengadaan. Sebagai catatan faktor tersebut hanya akan diterapkan pada suku cadang yang

memiliki peluang 50% untuk mengetahui kecenderungan suku cadang tersebut. Hasil selengkapnya ditunjukkan oleh Tabel 5.5., Tabel 5.6., dan Tabel 5.7.

Tabel 5.5. Hasil Integrasi *Clustering* Bulan Januari

Januari		
Material Code	Probabilitas	Kluster
39	50.0%	2
45	50.0%	1
46	50.0%	1
47	66.7%	1
67	50.0%	1
71	50.0%	0
74	50.0%	1
82	83.3%	2
83	100.0%	1
152	50.0%	0

Dengan penerapan faktor yang sebelumnya dijelaskan maka pada bulan Januari untuk material 71 dan 152 tidak direkomendasikan untuk dilakukan pengadaan karena material tersebut memiliki peluang 50% serta berada pada kluster pertama yang menjadi faktor yang melemahkan.

Tabel 5.6. Hasil Integrasi *Clustering* Bulan Februari

Februari		
Material Code	Probabilitas	Kluster
47	66.7%	1
49	66.7%	0
50	50.0%	0
52	50.0%	0
64	50.0%	1
67	50.0%	1
69	50.0%	2
74	66.7%	1
78	50.0%	1
79	50.0%	0

82	66.7%	2
83	66.7%	1
85	66.7%	1
86	66.7%	1
112	50.0%	0
194	50.0%	0
197	50.0%	0
199	50.0%	0
212	50.0%	0
323	50.0%	2
328	66.7%	1
329	50.0%	0

Dengan penerapan faktor yang sebelumnya dijelaskan maka pada bulan Februari untuk material 50, 52, 79, 112, 194, 197, 199, 212, dan 329 tidak direkomendasikan untuk dilakukan pengadaan karena material tersebut memiliki peluang 50% serta berada pada kluster pertama yang menjadi faktor yang melemahkan. Di samping itu, untuk material 49 merupakan material yang tergolong dalam kluster pertama atau memiliki sifat *non moving* dan *low demand* juga akan dihilangkan.

Tabel 5.7. Hasil *Clustering* Integrasi Bulan Maret

Maret		
Material Code	Probabilitas	Kluster
32	50.0%	2
36	66.7%	1
39	100.0%	2
45	50.0%	1
46	66.7%	1
47	83.3%	1
51	66.7%	0
52	83.3%	0
69	50.0%	2
71	50.0%	0
78	50.0%	1
82	66.7%	2
83	66.7%	1
86	66.7%	1
114	66.7%	0

196	50.0%	0
197	66.7%	0
326	50.0%	1
328	83.3%	1
334	50.0%	0

Dengan penerapan faktor yang sebelumnya dijelaskan maka pada bulan Maret untuk material 71, 196, dan 334 tidak direkomendasikan untuk dilakukan pengadaan karena material tersebut memiliki peluang 50% serta berada pada kluster pertama yang menjadi faktor yang melemahkan. Di samping itu, untuk material 51, 52, 114, dan 197 merupakan material yang tergolong dalam kluster pertama atau memiliki sifat *non moving* dan *low demand* juga akan dihilangkan.

Terakhir, analisis AR dapat ditambahkan untuk mencari tahu kombinasi material yang sering untuk muncul. Hal ini bermanfaat dalam menghemat biaya logistik ataupun biaya pemesanan karena pemberian diskon untuk jumlah pesanan tertentu dapat saja berlaku. Sebagai contoh material 39 dapat dipesan bersamaan dengan material 46 dan 47 dengan tingkat *confidence* dapat dilihat pada Tabel 5.1. Hasil kombinasi selengkapnya ditunjukkan oleh Tabel 5.8., Tabel 5.9., dan Tabel 5.10.

Tabel 5.8. Hasil Integrasi AR Bulan Januari

Januari			
Material Code	Probabilitas	Kluster	Material Penyerta
39	50.0%	2	46,47
45	50.0%	1	46,47,82,39
46	50.0%	1	47,45,39
47	66.7%	1	46,83
67	50.0%	1	-
74	50.0%	1	-
82	83.3%	2	83,78
83	100.0%	1	82

Tabel 5.9. Hasil Integrasi AR Bulan Februari

Februari			
Material Code	Probabilitas	Kluster	Material Penyerta
47	66.7%	1	46,83
64	50.0%	1	-
67	50.0%	1	-
69	50.0%	2	46,82
74	66.7%	1	-
78	50.0%	1	82
82	66.7%	2	83,78
83	66.7%	1	82
85	66.7%	1	82,83
86	66.7%	1	83
323	50.0%	2	-
328	66.7%	1	47

Tabel 5.10. Hasil Integrasi AR Bulan Maret

Maret			
Material Code	Probabilitas	Kluster	Material Penyerta
32	50.0%	2	-
36	66.7%	1	-
39	100.0%	2	46,47
45	50.0%	1	46,47,82,39
46	66.7%	1	47,45,39
47	83.3%	1	46,83
69	50.0%	2	46,82
78	50.0%	1	82
82	66.7%	2	83,78
83	66.7%	1	82
86	66.7%	1	83
326	50.0%	1	-
328	83.3%	1	47

5.5. Diskusi

Penelitian yang dilakukan dalam laporan ini dimaksudkan untuk mencari tahu pengetahuan apa saja yang bisa didapatkan terkait dengan data kebutuhan suku cadang.

Oleh karena itu, digunakan 3 metode terkait *clustering*, *forecasting*, dan *association rule*. Dengan penggunaan ketiga metode tersebut dapat diketahui banyaknya kebutuhan material dan material yang berkemungkinan untuk muncul serta faktor pendukung. Namun, dalam pengolahan data yang dilakukan muncul potensi pengembangan lain untuk mengetahui lebih banyak pengetahuan dari data yang ada.

Kemungkinan pengembangan pertama terkait dengan metode *clustering*. Dalam penelitian ini, *clustering* digunakan sebagai faktor pemberat atau peringan untuk material yang ada karena data yang ada akan terbagi ke dalam 3 kluster dengan sifat *low demand*, *middle demand*, dan *high demand*. Dengan adanya pembagian ini, *user* dapat menggunakan hasil *clustering* sebagai bahan pertimbangan dalam proses pengadaan. Namun, hal tersebut menimbulkan adanya subjektifitas dalam pengambilan keputusan. Oleh karena itu, hasil dari *clustering* tersebut dapat dikembangkan kembali untuk mengurangi subjektifitas tersebut. Pengembangan dapat berupa analisis lebih dalam terhadap material yang tergabung dalam kluster *middle demand* dan *high demand* karena kedua kluster tersebut menjadi faktor pemberat. Material tersebut dapat dianalisis apakah bersifat *fast moving*, *slow moving*, atau *no moving*. Adanya analisis tersebut memungkinkan untuk mengetahui frekuensi pengadaan yang dilakukan. Dengan mengetahui sifat tersebut kelompok dari material akan menjadi semakin jelas serta dapat mempengaruhi hasil peramalan secara positif. Hal ini dikarenakan data yang digunakan dalam peramalan berbentuk *aggregate*, sehingga sifat dari setiap material bercampur yang mungkin mengakibatkan turunnya akurasi pada saat proses peramalan. Sebagai contoh jika sebuah material memiliki sifat *no moving* atau pengadaan hanya dilakukan sekali dalam 72 bulan hal ini akan membuat keseluruhan data menjadi bias. Oleh karena itu, pengelompokan lanjutan perlu untuk dilakukan agar akurasi peramalan dapat meningkat. Analisis lain yang dapat dilakukan yaitu kuantifikasi dari *clustering* yang dihasilkan. Dengan dilakukannya hal tersebut maka tingkat prioritas dari sebuah material dapat diketahui serta jika digabungkan dengan metode lain dapat diketahui bobot dari faktor *clustering* ini.

Pengembangan lain yang dapat dilakukan berkaitan dengan hasil metode *forecasting*. Berdasarkan Gambar 4.14 diketahui bahwa data tersebut terdapat puncak pada setiap tahunnya. Namun, dalam waktu 72 bulan puncak tersebut tidak beraturan atau tidak menunjukkan adanya pola siklus dalam data tersebut. Hal ini dapat dilakukan analisis lebih dalam untuk mengetahui apakah benar data tersebut tidak memiliki pola siklus karena jika pada tahun 2013 salah satu puncak dihilangkan maka muncul indikasi bahwa data tersebut memiliki pola siklus. Hal ini dapat saja diakibatkan oleh data *noise* atau hal lain yang dapat mempengaruhi. Selain itu penambahan data yang digunakan dapat dilakukan untuk meningkatkan akurasi dari peramalan karena *machine learning* membutuhkan data *training* yang cukup banyak untuk proses pembelajaran.



BAB VI

KESIMPULAN DAN SARAN

6.1. Kesimpulan

Berdasarkan pengolahan data dan pembahasan yang telah dilakukan pada BAB IV dan BAB V dapat ditarik kesimpulan sebagai berikut:

1. Kluster yang terbentuk terdiri dari 3 kelompok dengan pembagian *low demand – non moving*, *middle demand – slow moving*, dan *high demand – fast moving*. Pembagian ini berdasarkan dari nilai rata-rata setiap kluster atau pada *low demand* dan *non moving* memiliki nilai rata-rata sebesar 67.54 dan 4, *middle demand* dan *slow moving* memiliki rata-rata sebesar 1281.17 dan 24, dan *high demand* dan *fast moving* memiliki rata-rata sebesar 2474 dan 30. Selain itu kluster *low demand* memiliki anggota 327 material, *middle demand* memiliki anggota 18 material, dan *high demand* memiliki anggota 5 material.
2. Prediksi untuk kebutuhan pengadaan untuk bulan Januari sebesar 1241, bulan Februari sebesar 201, dan bulan Maret sebesar 500. Di samping prediksi jumlah kebutuhan juga dilakukan prediksi mengenai material apa saja yang berkemungkinan untuk muncul dengan nilai peluang minimal 50%. Pada bulan Januari material 39, 45, 46, 47, 67, 74, 82, dan 83 berkemungkinan untuk muncul. Pada bulan Februari material 47, 64, 67, 69, 74, 78, 82, 83, 85, 86, 323, dan 328 berkemungkinan untuk muncul. Pada bulan Maret material 32, 36, 39, 45, 46, 47, 69, 78, 82, 83, 86, 326, dan 328 berkemungkinan untuk muncul. Namun, karena material tersebut berada pada nilai tengah maka perlu untuk dilakukan analisis tambahan guna mengetahui kecenderungan terjadinya.

3. Aturan asosiasi atau kombinasi material yang didapatkan dari analisis AR berjumlah 24 kombinasi. Kombinasi ini memiliki nilai minimal *support* sebesar 10% dan minimal *confidence* sebesar 50%.
4. Integrasi antara ketiga metode tersebut dapat dilakukan untuk mengatasi kekurangan yang ada. Hasil dari peramalan menunjukkan tingkat kebutuhan suku cadang untuk 3 bulan mendatang, namun hasil tersebut masih bersifat *aggregate* atau gabungan dari semua material yang ada. Oleh karena itu, untuk memprediksi ke tingkat individual dapat digunakan metode peluang. Hasil *clustering* ditambahkan untuk mengetahui karakteristik dari material yang ada. Sehingga, tingkat prioritas dapat ditentukan serta material yang memiliki peluang 50% dapat diketahui kecenderungan terjadinya. Dengan integrasi *clustering* diketahui terdapat beberapa material yang tidak direkomendasikan untuk dilakukan pengadaan karena dinilai cenderung tidak akan terjadi. Aturan yang terbentuk dari analisis AR dapat ditambahkan untuk mengetahui material penyerta atau kombinasi material yang cenderung akan terjadi. Dengan kata lain, hasil analisis AR juga menjadi faktor pemberat untuk suatu material akan muncul atau cenderung terjadi dalam suatu transaksi.

6.2. **Saran**

Saran yang dapat diberikan adalah sebagai berikut:

1. *Clustering* dalam penelitian ini digunakan sebagai faktor pendukung dalam penentuan pengadaan suatu material. Namun, penggunaan tersebut masih memiliki nilai subjektifitas. Oleh karena itu, penelitian selanjutnya dapat mencari tahu kuantifikasi dari *clustering* untuk mengurangi subjektifitas dalam penentuan keputusan serta menjadi bobot *clustering* jika digabungkan dengan metode lain.
2. Hasil dari penelitian ini memiliki kekurangan dalam mengetahui perkiraan jumlah pengadaan untuk setiap material yang ada. Oleh karena itu penelitian selanjutnya dapat memberikan analisis tambahan untuk mengetahui hal tersebut.

3. Penelitian selanjutnya juga dapat berfokus pada kluster *middle* dan *high demand* untuk mengetahui lebih lanjut karakteristik dari material yang tergabung dalam kluster tersebut. Contoh analisis yang dapat dilakukan adalah untuk mencari tahu apakah material yang tergabung bersifat *fast moving*, *slow moving*, atau *no moving* material. Analisis juga dapat dilakukan pada kelompok data yang bersifat *slow moving* atau *fast moving* jika analisis mengenai sifat tersebut telah dilakukan.



DAFTAR PUSTAKA

- Andayani, S. (2007). Pembentukan cluster dalam Knowledge Discovery in Database dengan Algoritma K-Means. *SEMNAS Matematika Dan Pendidikan Matematika 2007*.
- Aprianti, W., Hafizd, K. A., & Rizani, M. R. (2017). Implementasi Association Rules dengan Algoritma Apriori pada Dataset Kemiskinan. *Limits: Journal of Mathematics and Its Applications*, 14(2), 57. <https://doi.org/10.12962/limits.v14i2.2933>
- Bacchetti, A., & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, 40(6), 722–737. <https://doi.org/10.1016/j.omega.2011.06.008>
- Biegel, J. E. (1999). *Pengendalian Produksi Suatu Pendekatan Kuantitatif*. Jakarta: Akademika Presindo.
- Buffa, E. S., & Sarin, R. K. (1996). *Manajemen Operasi/ Produksi Modern Jilid 1 -8/E*. Jakarta: Binarupa Aksara.
- Danukusumo, K. P. (2017). *Implementasi Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Citra Candi Berbasis Gpu*.
- Davies, & Beynon, P. (2004). *Database Systems Third Edition*.
- Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). An Introduction to Classification and Clustering (pp. 1–13). <https://doi.org/10.1002/9780470977811.ch1>
- Fauzy, M., Saleh W, K. R., & Asror, I. (2016). Penerapan Metode Association Rule Menggunakan Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung. *Jurnal Ilmiah Teknologi Informasi Terapan*, II(2), 221–227.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Journal of Japan Society for Fuzzy Theory and Systems*, 9(6), 851–860. https://doi.org/10.3156/jfuzzy.9.6_851
- Feng, Z., Bing, C., Shouquan, W., Gaolun, C., & Weibo, G. (2019). Research on Consumption Law Prediction of aircraft spares based on Holt-Winters. *Journal of Physics: Conference Series*, 1213(5). <https://doi.org/10.1088/1742-6596/1213/5/052017>
- Gamberini, R., Lolli, F., Rimini, B., & Sgarbossa, F. (2010). Forecasting of Sporadic Demand Patterns with Seasonality and Trend Components: An Empirical Comparison between Holt-Winters and (S)ARIMA Methods. *Mathematical Problems in Engineering*, 2010, 579010. <https://doi.org/10.1155/2010/579010>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Hall, C., Lynch, P., Michael, E. J., & Mitchell, R. (2007). Micro-clusters and networks: The growth of tourism. *Micro-Clusters and Networks: The Growth of Tourism*, 141–152.
- Han, J., & Kamber, M. (2006). *Data Mining : Concept and Techniques Second Edition*. Morgan Kaufmann Publishers.

- Harrington, H. J. (1991). Improving business processes. *The TQM Magazine*, 3(1).
<https://doi.org/10.1108/eb059514>
- Heizer, J., & Render, B. (2005). *Operations Management - Manajemen Operasi (Edisi Ketujuh)*. Jakarta: Salemba Empat.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Linoff, G., & Berry, M. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of Machine Learning*.
- Nasari, F., Jhony, C., & Sianturi, M. (2016). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat, 108–119.
- Ningrat, D. R., Maruddani, D. A. I., & Wuryandari, T. (2016). Analisis cluster dengan algoritma K-Means dan Fuzzy C-Means clustering untuk pengelompokan data obligasi korporasi. *None*, 5(4), 641–650.
- Nugraha, G. S., & Hairani, H. (2018). Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode K-Means. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 13–23.
- Pane, S. F., & Saputra, Y. A. (2020). *Big Data: Classification Behavior Menggunakan Python*. Bandung: CV. Kreatif Industri Nusantara.
- Pramudiono, I. (2003). Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data. *Kuliah Umum Ilmu Komputer.Com*, 1–4.
- Rerung, R. R. (2018). Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk. *Jurnal Teknologi Rekayasa*, 3(1), 89.
<https://doi.org/10.31544/jtera.v3.i1.2018.89-98>
- Rofiqo, N., Windarto, A. P., & Hartama, D. (2018). Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 2(1), 216–223.
<https://doi.org/10.30865/komik.v2i1.929>
- Suwarningsih, W. (2008). Aplikasi Data Mining dengan Menggunakan Teknik ARM untuk Pengolahan Informasi Rendemen Obat. *INKOM: Jurnal Informatika, Sistem Kendali, Dan Komputer*, 2(2), 1–5. Retrieved from
<https://www.neliti.com/id/publications/66970/aplikasi-data-mining-dengan-menggunakan-teknik-arm-untuk-pengolahan-informasi-re>
- Taylor, J. W. (2004). Smooth Transition Exponential Smoothing. *Journal of Forecasting*, 23(6), 385–404. <https://doi.org/https://doi.org/10.1002/for.918>
- Zanuardi, A., & Suprayitno, H. (2018). Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan Knowledge Discovery in Database. *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, 2.
<https://doi.org/10.12962/j26151847.v2i1.3767>

LAMPIRAN

A - Lampiran 1: Anggota Kluster Pertama

Material Code	PO Quantity	Freq	Clusters
1	20	2	0
2	24	3	0
3	50	1	0
4	24	3	0
5	121	9	0
6	32	4	0
7	25	1	0
8	4	2	0
9	26	2	0
10	74	1	0
11	2	1	0
12	34	4	0
13	144	6	0
14	233	14	0
15	80	6	0
16	205	6	0
17	72	6	0
18	2	1	0
19	1	1	0
20	50	1	0
21	12	2	0
22	58	2	0
23	2	1	0
24	16	2	0
25	3	1	0
26	50	1	0
27	18	1	0
28	2	1	0
29	4	2	0
30	10	1	0
31	6	1	0
38	65	3	0
42	80	3	0
43	126	8	0
49	605	26	0
50	418	21	0

51	304	23	0
52	263	26	0
54	5	4	0
55	9	3	0
56	2	2	0
57	5	4	0
58	2	1	0
60	89	6	0
62	194	9	0
65	499	6	0
71	621	28	0
72	25	2	0
73	12	3	0
75	388	16	0
76	47	4	0
79	186	18	0
81	85	6	0
88	368	16	0
90	207	9	0
91	15	3	0
92	143	13	0
94	42	4	0
95	94	14	0
96	7	2	0
97	19	4	0
98	8	5	0
99	6	1	0
100	8	2	0
102	30	2	0
103	11	2	0
104	23	4	0
106	8	2	0
108	50	4	0
109	52	5	0
110	166	9	0
111	178	8	0
112	220	13	0
113	470	13	0
114	277	17	0
115	110	14	0
116	55	7	0
117	138	8	0
118	50	5	0
119	88	9	0
120	2	1	0
121	9	3	0

123	100	1	0
124	10	2	0
125	3	1	0
126	25	2	0
127	24	2	0
128	96	5	0
129	98	6	0
130	50	3	0
131	6	2	0
133	6	2	0
135	99	3	0
137	4	1	0
138	26	6	0
139	5	3	0
141	2	1	0
142	53	5	0
145	8	1	0
146	17	3	0
148	25	5	0
150	345	10	0
152	437	11	0
154	226	10	0
156	16	1	0
157	121	6	0
158	6	3	0
159	3	1	0
160	6	2	0
161	4	1	0
162	17	4	0
164	2	1	0
165	6	3	0
166	34	3	0
168	24	3	0
169	200	6	0
170	20	4	0
171	1	1	0
172	98	11	0
173	2	1	0
174	5	4	0
175	10	4	0
176	30	3	0
177	11	2	0
178	10	1	0
179	60	6	0
180	6	2	0
181	8	2	0

182	40	5	0
183	7	3	0
184	13	3	0
185	75	5	0
187	33	5	0
188	18	4	0
189	9	2	0
190	70	4	0
191	218	6	0
192	79	4	0
193	14	3	0
194	231	11	0
196	488	17	0
197	492	22	0
198	227	10	0
199	165	9	0
200	63	5	0
201	8	2	0
202	4	2	0
203	22	4	0
204	4	2	0
205	2	1	0
206	6	3	0
207	5	3	0
208	8	1	0
209	11	3	0
210	129	5	0
211	37	4	0
212	139	11	0
213	201	11	0
214	56	7	0
215	218	11	0
216	164	6	0
217	144	7	0
218	1	1	0
219	25	2	0
220	53	7	0
221	2	2	0
222	2	1	0
223	4	2	0
224	58	1	0
225	48	4	0
226	4	2	0
227	10	2	0
228	2	1	0
229	1	1	0

230	6	3	0
231	8	4	0
232	141	6	0
233	28	3	0
234	16	2	0
236	12	3	0
237	47	4	0
239	19	3	0
240	10	1	0
241	10	1	0
242	8	1	0
243	8	2	0
244	20	1	0
245	34	3	0
246	6	1	0
247	32	5	0
249	3	1	0
250	8	1	0
251	4	1	0
252	8	3	0
253	20	3	0
254	4	1	0
255	81	2	0
256	18	2	0
257	36	2	0
258	13	3	0
259	9	1	0
260	12	2	0
261	38	4	0
262	4	2	0
263	36	5	0
265	8	3	0
266	34	4	0
267	2	1	0
268	46	4	0
269	3	1	0
270	30	1	0
271	3	1	0
272	23	4	0
274	12	1	0
275	2	1	0
276	4	1	0
277	1	1	0
278	2	1	0
279	4	1	0
280	5	1	0

281	8	2	0
282	8	2	0
283	6	3	0
284	13	2	0
285	30	1	0
286	18	4	0
287	18	4	0
288	13	3	0
289	32	1	0
290	11	4	0
291	41	3	0
292	96	3	0
293	82	3	0
294	54	2	0
295	51	4	0
296	4	1	0
297	2	1	0
298	15	4	0
299	2	1	0
300	4	1	0
301	29	2	0
302	4	1	0
303	3	2	0
304	6	1	0
305	101	4	0
306	71	5	0
307	16	1	0
308	39	3	0
309	25	1	0
310	68	3	0
311	32	3	0
312	33	3	0
313	2	1	0
314	2	1	0
315	10	2	0
316	8	1	0
317	28	5	0
318	2	1	0
319	2	1	0
320	4	2	0
321	2	1	0
322	18	2	0
329	191	15	0
330	356	22	0
332	37	10	0
333	2	1	0

334	181	9	0
335	105	6	0
336	5	1	0
337	530	10	0
338	23	6	0
339	303	6	0
340	7	1	0
341	264	2	0
342	106	6	0
343	6	3	0
344	35	4	0
345	250	1	0
346	12	3	0
347	206	7	0
348	126	4	0
349	108	5	0
350	62	5	0
351	51	5	0
352	30	3	0
353	20	2	0
354	56	2	0
355	290	3	0
356	116	3	0
357	59	4	0
358	100	4	0
359	179	7	0
360	142	9	0
361	52	7	0
362	92	3	0
363	59	5	0
364	10	3	0
365	12	2	0
366	30	5	0
367	19	2	0
368	2	1	0
369	1	1	0
370	1	1	0
371	1	1	0
372	250	1	0
373	250	1	0
374	250	1	0
375	251	2	0
376	250	1	0
377	124	2	0
378	5	1	0
379	4	2	0

380	8	2	0
381	8	2	0
382	12	2	0
383	4	2	0
384	20	2	0
385	32	1	0
386	200	6	0
387	8	1	0
388	2	1	0
389	5	1	0
390	4	1	0
391	1	1	0
392	2	1	0
393	1	1	0
394	3	1	0

B - Lampiran 2: Anggota Kluster Kedua

Material Code	PO Quantity	Freq	Clusters
35	1576	18	1
36	1649	22	1
45	1255	32	1
46	1513	29	1
47	1077	34	1
59	1158	14	1
61	1586	15	1
63	1528	15	1
64	1464	19	1
66	1501	20	1
67	1011	21	1
74	789	25	1
78	1502	25	1
83	1794	38	1
85	832	24	1
86	845	27	1
326	1066	25	1
328	915	30	1

C - Lampiran 3: Anggota Kluster Ketiga

Material Code	PO Quantity	Freq	Clusters
---------------	-------------	------	----------

32	2047	26	2
39	2632	37	2
69	2576	27	2
82	2371	34	2
323	2744	24	2

