

**IMPLEMENTASI *DATA MINING* MENGGUNAKAN METODE
ASSOCIATION RULE PADA REGISTER USER KOMPAS.ID**

SKRIPSI

**OLEH:
AULIA GITA PRATIWI
NIM. 19610057**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**IMPLEMENTASI *DATA MINING* MENGGUNAKAN METODE
ASSOCIATION RULE PADA *REGISTER USER* KOMPAS.ID**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
AULIA GITA PRATIWI
NIM. 19610057**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**IMPLEMENTASI DATA MINING MENGGUNAKAN METODE
ASSOCIATION RULE PADA REGISTER USER KOMPAS.ID**

SKRIPSI

Oleh
Aulia Gita Pratiwi
NIM. 19610057

Telah Disetujui untuk Diuji
Malang, 8 Januari 2024

Dosen Pembimbing I



Angga Dwi Mulyanto, M.Si.
NIP. 19890813 201903 1 012

Dosen Pembimbing II



Erna Herawati, M.Pd.
NIP. 197607232 02321 2 006

Mengetahui,
Ketua Program Studi Matematika



Dwi Susanti, M.Sc.
NIP. 19741129 200012 2005

**IMPLEMENTASI DATA MINING MENGGUNAKAN METODE
ASSOCIATION RULE PADA REGISTER USER KOMPAS.ID**

SKRIPSI

**Oleh
Aulia Gita Pratiwi
NIM. 19610057**

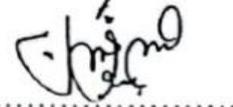
Telah Dipertahankan di Depan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal, 20 Februari 2024

Ketua Penguji : Abdul Aziz, M.Si.



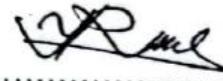
Anggota Penguji 1 : Ria Dhea Layla Nur Karisma, M.Si.



Anggota Penguji 2 : Angga Dwi Mulyanto, M.Si.



Anggota Penguji 3 : Erna Herawati, M.Pd



Mengetahui,
Ketua Program Studi Matematika




Elly Susanti, M.Sc.
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Aulia Gita Pratiwi
NIM : 19610057
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Implementasi *Data Mining* Menggunakan
Association Rule Pada *Register User* Kompas.id

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Februari 2024
Yang membuat pernyataan,

Aulia Gita Pratiwi
NIM. 19610057

MOTO

“The great aim of education is not knowledge, but action.”

PERSEMBAHAN

Skripsi ini penulis persembahkan kepada:

Eyang kong Diran tersayang, yang cintanya seluas lautan. Ayah dan Ibu tercinta Fauzi dan Heny yang selalu menerima bungsunya dengan apa adanya. Mas Deo, Abang *terjail*. Almh. Bu Kumolo, guru matematika sewaktu SMP yang membuat penulis mencintai matematika. Serta sahabat-sahabat penulis yang selalu mendengarkan dan membantu penulis.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah Swt. Sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Karya ini tidak akan diselesaikan dengan baik tanpa dukungan dan bantuan orang-orang yang telah membantu. Terima kasih penulis sampaikan kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Dr. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, M.Sc., selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim.
4. Angga Dwi Mulyanto, M.Si., selaku dosen pembimbing I, selaku dosen pembimbing I dan penguji II yang telah memberikan masukan, pengetahuan, kemudahan dan kebaikan yang luar biasa kepada penulis.
5. Erna Herawati, M.Pd., selaku dosen pembimbing II yang memberikan ilmu dan pelajaran kepada penulis
6. Abdul Aziz, M.Si., selaku ketua penguji dalam ujian skripsi yang telah memberikan saran yang bermanfaat bagi penulis
7. Ria Dhea Layla Nur Karisma, M.Si., selaku penguji I dalam ujian skripsi yang telah memberikan saran yang bermanfaat bagi penulis.
8. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim
9. Ayah Fauzi dan Ibu Heny selaku orang tua penulis serta seluruh keluarga yang senantiasa memberikan dukungan dan doa.
10. Seluruh mahasiswa Matematika Angkatan 2019 seperjuangan.

Malang, 20 Februari 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGANTAR	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
ABSTRAK	xiii
ABSTRACT	xiv
مستخلص البحث	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian	6
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah	7
1.6 Definisi Istilah	7
BAB II KAJIAN TEORI	9
2.1 <i>Data Mining</i>	9
2.1.1 Pengertian <i>Data Mining</i>	9
2.1.2 Pengelompokan <i>Data Mining</i>	12
2.1.3 Tahapan <i>Data Mining</i>	13
2.2 <i>Association Rule</i>	15
2.2.1 Pengertian <i>Association Rule</i>	15
2.2.2 Ukuran Kepercayaan <i>Rule</i>	17
2.3 Algoritma Apriori	28
2.4 Kajian Integrasi Topik dengan Al-Quran/Hadits	31
2.5 Kajian Topik Dengan Teori Pendukung	35
BAB III METODE PENELITIAN	37
3.1 Jenis Penelitian	37
3.2 Data dan Sumber Data	37
3.3 Teknik Pengumpulan Data	40
3.4 Instrumen Penelitian	40
3.5 Teknik Analisis Data	41
3.6 <i>Flowchart</i>	42
BAB IV HASIL DAN PEMBAHASAN	43
4.1 Filterisasi Data	43
4.2 Pengujian Minimum <i>Support</i> (α)	45
4.3 Pengujian Nilai Minimum <i>Confidence</i>	47
4.4 Pengujian Nilai <i>Lift</i>	49

4.5	Pembuatan Pola <i>Association Rule</i>	50
4.6	Kajian Integrasi Topik Dalam Islam.....	52
BAB V	KESIMPULAN	55
5.1	Kesimpulan	55
5.2	Saran	55
	DAFTAR PUSTAKA	57
	LAMPIRAN	60
	RIWAYAT HIDUP	61

DAFTAR TABEL

Tabel 2.1	Transaksi <i>Database</i>	19
Tabel 2.2	Himpunan Item-1 Dalam <i>Database</i>	21
Tabel 2.3	Himpunan Set-2 Dalam <i>Database</i>	22
Tabel 2.4	Himpunan Set-3 Dalam <i>Database</i>	23
Tabel 2.5	Himpunan Item-4 Dalam <i>Database</i>	23
Tabel 2.6	<i>Association Rule</i> Dengan 1-Item Konsekuensi 3 Himpunan Set	24
Tabel 2.7	<i>Association Rules</i> Dengan 2-Item Konsekuensi 3 Himpunan Item	25
Tabel 2.8	<i>Association Rule</i> Untuk {Opini, Polhuk}	25
Tabel 2.9	<i>Association Rules</i> Untuk {Internasional, Polhuk}	26
Tabel 2.10	<i>Association Rule</i> Untuk {Internasional, Ekonomi}	26
Tabel 2.11	<i>Association Rules</i> Untuk {Polhuk, Ekonomi}	27
Tabel 2.12	Frekuensi Himpunan Set-1 Pada <i>Dataset</i> Tabel 2.1.....	30
Tabel 2.13	Frekuensi Himpunan Item-2 Pada <i>Dataset</i> Di Tabel 2.1	30
Tabel 2.14	Frekuensi Himpunan Item-3 Pada <i>Dataset</i>	31
Tabel 3.1	Jumlah Artikel dan Opini Yang Dibaca <i>User</i>	39
Tabel 4.1	Filterisasi Organik <i>User</i>	43
Tabel 4.2	Frequent itemset	44
Tabel 4.3	Pemfilteran Minimum <i>Support</i>	45
Tabel 4.4	Pembuatan <i>Support</i>	45
Tabel 4.5	Pengujian Nilai Minimum <i>Confidence</i>	48
Tabel 4.6	Pembentukan <i>Association Rule</i> Berdasarkan Nilai <i>Confidence</i>	48
Tabel 4.7	Penentuan nilai <i>lift</i>	49

DAFTAR GAMBAR

Gambar 2.1	Gambaran AI, ML, DL, <i>Data Mining</i> , dan <i>Computer Science</i>	11
Gambar 2.2	<i>Knowledge Discovery in Database (KDD)</i>	14
Gambar 3.1	<i>Pie Chart</i> Tipe User Kompas.id	38
Gambar 3.2	<i>Flowchart</i>	42
Gambar 4.1	Diagram <i>Association Rule Register User</i> Kompas.id.....	50

ABSTRAK

Pratiwi, Aulia Gita. 2023. **Implementasi *Data Mining* Menggunakan *Association Rule* Pada Register User Kompas.id**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: I) Angga Dwi Mulyanto, M.Si, II) Erna Herawati, M.Pd.

Kata Kunci: *Data Mining*, Algoritma Apriori, *Association Rule*, Artikel, Berita.

Data mining adalah metode analisis data baru yang mengekstrak informasi berharga dari data besar untuk digunakan oleh manusia. Penelitian ini bertujuan untuk mengetahui *association rule* artikel/berita yang dihasilkan dari penerapan analisis *market basket analysis* dengan menggunakan algoritma apriori, dengan mempertimbangkan 3 *top category* jenis artikel/berita yang sering dibaca *register user* Kompas.id. Pendekatan penelitian yaitu penerapan data mining dengan algoritma apriori yang dapat menemukan *frequent itemset mining* dan *association rule* (pola asosiasi). Nilai *minimum support* sebesar 0.021 atau 2%, nilai *confidence* sebesar 0.0541 serta nilai *lift* ≥ 1 . Hasil penelitian menunjukkan bahwa hasil pola *association rule* berdasarkan nilai *support* tertinggi terdapat pada artikel {Humaniora} \Rightarrow {Opini}. Hampir semua nilai *confidence* bernilai 1, maka dapat disimpulkan bahwa setiap asosiasi antara LHS dan RHS memiliki korelasi yang kuat. Nilai *lift* pada seluruh asosiasi bernilai 1, artinya terdapat korelasi yang kuat pada seluruh asosiasi yang ditemukan. Penelitian ini menyimpulkan bahwa pada metode *association rule* dan algoritma apriori, tingkat *support* tinggi menunjukkan adanya hubungan yang signifikan antara dua artikel dan atau lebih, tingkat *confidence* yang tinggi mengindikasikan bahwa ketika satu jenis artikel/berita muncul, jenis artikel lainnya juga cenderung muncul. *Lift* tinggi menunjukkan terdapat ketergantungan yang kuat antara tiap-tiap *rule* yang ada.

ABSTRACT

Pratiwi, Aulia Gita, 2023: **Implementation of Data Mining Using Association Rule on Kompas.id User Registration**. Thesis. Department of Mathematics, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Advisors: I) Angga Dwi Mulyanto, M.Si, II) Erna Herawati, M.Pd.

Keywords: Data Mining, Apriori, Association Rule, Articles, News.

Data mining is a new data analysis method that extracts valuable information from large datasets for human use. This research aims to determine the association rules of articles/news generated from the application of market basket analysis using the Apriori algorithm, considering the top 3 categories of articles/news that are frequently read by Kompas.id user registrations. The research approach involves the application of data mining with the Apriori algorithm, which can discover frequent itemset mining and association rules (association patterns). The minimum support is 0.021 or 2.1% and confidence values in this study are 0.0541 or 5.41%, and the lift value is ≥ 1 . The research results show that the highest support-based association rules are found in articles {Humanities} \Rightarrow {Opinion}. Almost all confidence values are equal to 1, indicating that there is a strong correlation between every association between LHS and RHS. The lift values for all associations are 1, meaning there is a strong correlation in all the associations found. This research concludes that in the association rule method, a high level of support indicates a significant relationship between two or more articles, and a high level of confidence indicates that when one type of article/news appears, other types of articles are also likely to appear. High lift values indicate a strong dependency between each existing rule.

مستخلص البحث

براتبوي، أوليا جيتا. ٢٠٢٣. تنفيذ استخراج البيانات باستخدام قواعد الارتباط في المسترزمين *kompas id*. البحث الجامعي فليسليم لرياضيات، كلية العلوم والتكنولوجيا جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. الشرف: (الأول أنجا دوي موليانتو، الماجستير) المشرف الثاني أرنا هيراواتي الماجستير.

كلمة الرئيسية: التنقيب في البيانات، خوارزمية أبروري، قاعدة الارتباط، الأخبار.

التنقيب في البيانات هو طريقة جديدة في البيانات استخراج معلومات قيمة من البيانات الضخمة للاستخدامها من قبل البشر. هدف هذا البحث إلى تحديد قواعد ارتباط للمقالات/الأخبار التي تم إنشاؤها من تطبيق تحليل سلة السوق باستخدام خوارزمية أبروري من خلال النظر في فن فئات للمقالات/الأخبار التي يتم قراءتها بشكل متكرر من قبل تسجيلات مستخدمي *Kompas.id*. تشمل الطريقة البحثية تطبيق تنقيب البيانات بخوارزمية أبروري التي يمكنها اكتشاف تنقيب المجموعة المتكررة وقواعد الرابطة (أنماط الرابطة). قيم الدعم والثقة الدنيا في هذه الدراسة هي 2.1% أو 0.021، وقيمة الرفع هي $I \leq$. تظهر نتائج البحث أن قواعد الرابطة الأعلى قائمة على أساس الدعم توجد في المقالات {العلوم الإنسانية} \leq {الرأي}. تكاد قيم الثقة تكون متساوية I ، مما يشير إلى وجود ترابط قوي بين كل رابطة بين LHS و RHS . قيم الرفع لجميع الربطات هي I ، مما يعني وجود ترابط قوي في جميع الربطات المكتشفة. يستنتج هذا البحث أنه في طريقة قاعدة الرابطة، يشير مستوى عالٍ من الدعم إلى وجود علاقة هامة بين نوعين أو أكثر من المقالات، ويشير مستوى عالٍ من الثقة إلى أنه عند ظهور نوع واحد من المقالات/الأخبار، فإنه من المرجح أيضًا أن تظهر أنواع أخرى من المقالات. تشير قيم الرفع العالية إلى وجود تبعية قوية بين كل قاعدة موجودة

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining adalah metode analisis data baru yang mengekstrak informasi berharga dari data besar untuk digunakan oleh manusia. *Data mining* melibatkan analisis data untuk menemukan pola atau hubungan antar data yang sebelumnya tidak diketahui, proses ini juga dikenal sebagai *Knowledge Discovery in Database* (KDD) (Fayyad dkk., 1996). *Data mining* berfungsi untuk mengambil informasi dan data tertentu yang dapat diolah untuk menjadi struktur yang dapat dipahami manusia. Hal ini melibatkan sejumlah tahapan, seperti basis data dan manajemen data, pra-pemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi dan *online updating*. *Data mining* merupakan teknik yang relatif mudah dan cepat untuk menemukan relasi dan atau pola secara cepat. Secara umum tugas *data mining* dibagi menjadi enam kelompok yaitu, klasifikasi, klusterisasi, regresi, deteksi *anomaly*, *association rule learning* dan *summarization* (Fayyad, dkk., 1996). Metode yang digunakan pada *data mining* adalah irisan dari AI (*artificial intelligence*), statistika, *machine learning*, dan *database systems* (Kesavaraj & Sukumaran, 2013).

Saat ini *data mining* banyak diimplementasikan terhadap berbagai bidang seperti seni, teknik, politik, bisnis, jurnalisme dan lain-lain. Dalam bidang *marketing* dan bisnis, *association rule* merupakan salah satu konsep menarik dalam bidang *data mining* yang bertujuan untuk mencari hubungan atau asosiasi dalam

data. Teknik ini mengidentifikasi pola pembelian produk yang sering dibeli bersama oleh konsumen dengan tujuan untuk menemukan wawasan sebagai pengambilan keputusan bisnis. Hal ini melibatkan analisis data yang besar untuk melihat pola pembelian umum dan mengungkap hubungan antar produk atau kategori berbeda. Hal ini tujuannya adalah untuk mengoptimalkan strategi pemasaran, meningkatkan penjualan dan kepuasan pelanggan secara keseluruhan *lift* (Rerung, 2018).

Dijelaskan pula pada surah Al-Hujurat surat ke 6 (Kemenag RI, 2023):

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنْ جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَنْ تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْبِحُوا
عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ ﴿٦﴾

Artinya: “Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti, agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu.”

Menurut tafsir dalam surah tersebut dijelaskan tentang haramnya mengambil berita dari orang fasik tanpa melakukan klarifikasi (tabayyun) kebenarannya. Pengambilan berita dari orang fasik dikhawatirkan akan membahayakan bagi orang lain. Dalam istilah ushul fiqh, ayat ini juga menunjukkan larangan untuk mengikuti tata cara orang-orang fasik. (Kemenag RI, 2023).

Dijelaskan pada surat An-Nur ayat 19:

إِنَّ الَّذِينَ يُجِبُونَ أَنْ تَشِيَعَ الْفَاحِشَةُ فِي الَّذِينَ آمَنُوا لَهُمْ عَذَابٌ أَلِيمٌ فِي الدُّنْيَا وَالْآخِرَةِ
وَاللَّهُ يَعْلَمُ وَأَنْتُمْ لَا تَعْلَمُونَ ﴿١٩﴾

Artinya: “Sesungguhnya orang-orang yang ingin agar (berita) perbuatan yang amat keji itu tersiar di kalangan orang-orang yang beriman, bagi mereka azab yang pedih di dunia dan di akhirat. Dan Allah mengetahui, sedang, kamu tidak mengetahui.”

Menurut tafsir Ibnu Katsir pada surat An-Nur ayat 19, Sekertaris Dewan Hisbah Persatuan Islam (Persis), KH Zae Nandang menjelaskan, Allah SWT akan memberikan ancaman di dunia dan di akhirat kepada seseorang yang menyebarkan kabar bohong di Tengah kaum mukmin. Kaum mukmin harus cerdas dalam mencerna kabar atau berita, apakah memang kebenarannya seperti itu atau tidak. Maka dari itu, bila seorang mukmin tahu akan berita bohong yang disebarkan, maka sebaiknya tidak meneruskan kabar buruk tersebut kepada mukmin lainnya. (Kementerian Agama RI, 2022)

Penemuan *association rule* dibagi menjadi 2 fase: deteksi himpunan *item* yang sering muncul dan generasi *association rule*. *Association rule* memiliki ukuran asosiasi sehingga *rule* yang dihasilkan tidak akan menghasilkan *rule* yang terlalu banyak. Terdapat tiga metrik untuk membantu mengukur *association rule*, yaitu *support*, *confidence* dan *lift* (Rerung, 2018).

Mining dari himpunan itemset yang sering muncul adalah tahap penting dalam *association mining* yang menemukan himpunan item yang sering muncul dalam *database* transaksi. *Mining* dari himpunan *itemset* merupakan inti dalam banyak tugas *data mining* yang mencoba mencari pola menarik dari *dataset*. Algoritma Apriori merupakan bagian dari aturan asosiasi pada *Data Mining*. Algoritma apriori digunakan untuk menemukan pencarian *frequent item set mining* dan *association rule learning* (Kumbhare, 2014). Perhitungan analisis pada apriori adalah dengan menemukan nilai minimum dari *support* dan nilai minimum dari *confidence* sehingga akan menghasilkan *association rule* (Santoso, 2021). Dalam penelitian ini, akan digunakan istilah “*association rule*” untuk mempermudah penelitian.

Telah dilakukan penelitian berupa jurnal hasil diantaranya adalah bahwa *market basket network* berbasis analisis keranjang belanja dan memperluas hubungan dari keranjang belanja ke seluruh produk dalam jaringan. *Market basket analysis* memberikan informasi produk yang berbeda dari data transaksi yang sama dan perlu dihubungkan dengan efektivitas sistem rekomendasi (Kim dkk, 2012). Selain itu, menggunakan teknik *data mining* dapat menghasilkan strategi retensi pelanggan berdasarkan tingkat loyalitas mereka dan menemukan aturan asosiasi penting berdasarkan tingkat *confidence* yang tinggi menggunakan algoritma Apriori sehingga dapat meningkatkan kinerja bisnis (Silva dkk, 2019).

Pada penelitian lain ditemukan bahwa *association rule* dapat mengetahui kondisi utama kelainan pada berbagai jalur produksi, dan apakah ada *rule* terkait untuk kegagalan produk dan kondisi abnormal. *Association rule* abnormal pada produk yang dianalisis dapat memberikan informasi yang sistematis, ilmiah, dan kuantitatif kepada perusahaan melalui analisis data yang terakumulasi di perusahaan (Lin dkk, 2019). Transaksi dapat menemukan item apa yang muncul bersamaan sehingga dapat memberikan *insight* untuk pemasaran suatu produk dengan cara yang efektif, cara mengklasifikasikan produk pada *layout* toko atau kemasan produk atau item mana yang harus ditawarkan sebagai diskon untuk produk lain. Salah satu metode yang digunakan adalah algoritma apriori. Algoritma apriori adalah metode inovatif untuk menentukan *association rule* dari data besar mengikuti hasil implikasi yang terdiri dari lebih dari satu item. Dengan menemukan *rules* akan didapatkan *confidence* dan *support* dari semua produk dengan semua kombinasi *itemset* yang muncul (Husna dkk, 2019). Hasil penelitian menunjukkan juga bahwa algoritma Apriori dapat digunakan mengidentifikasi pola

pembelian konsumen dan menghasilkan aturan asosiasi yang berguna dalam meningkatkan penjualan dan keuntungan toko retail (Dewi dkk, 2022).

PT. Kompas Media Nusantara merupakan perusahaan yang bergerak di bidang media massa yang menerbitkan Harian Kompas dan Kompas.id. Kompas.id diluncurkan pada 2 Februari 2017 oleh Yayasan Bentara Rakyat (Kompas, 2018). Terdapat 2 jenis *users* pada Kompas.id, diantaranya *subscriber user* dan *register user*. *Subscriber user* adalah *user* yang telah membuat akun di Kompas.id namun tidak berlangganan, sedangkan *register user* adalah *user* yang telah membuat akun di Kompas.id dan berlangganan. Pada penelitian kali ini, akan terfokus pada *register user* saja karena tujuan pada penelitian ini adalah untuk menganalisis perilaku *register user* agar mereka terus berlangganan di Kompas.id dan memberikan *insight* untuk pemasaran artikel/berita dengan cara yang efektif, serta pengklasifikasian artikel/berita sebagai pada *layout website* untuk pembaca (Kompas, 2018).

Data *register user* yang ada pada *database* Kompas.id sudah terdokumentasi dengan baik, namun belum dimanfaatkan lebih maksimal untuk mengembangkan strategi pemasaran. Pada *website* Kompas.id terdapat beberapa rubrik, diantaranya politik & hukum (polhuk), ekonomi, opini, humaniora, riset, nusantara, metro, internasional, tokoh, olahraga, gaya hidup, multimedia, jelajah, video, fotografi, muda, hiburan, sastra, investigasi, buku, English, perjalanan, dan di balik berita (Redaksi, 2017). Masing-masing artikel/berita tersebut memiliki peminatnya masing-masing. Pemanfaatan data dapat digunakan menjadi rekomendasi bisnis dan dukungan fasilitas dengan memanfaatkan teknologi, hal inilah yang

menjadikan dasar penulis untuk melakukan penelitian *market basket analysis* menggunakan algoritma apriori sebagai penjualan artikel/berita di Kompas.id.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas dapat diambil rumusan masalah yaitu, bagaimana pola *association rule* artikel/berita yang dihasilkan dari penerapan analisis market basket analysis dengan menggunakan algoritma apriori?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini adalah untuk mengetahui *association rule* artikel/berita yang dihasilkan dari penerapan analisis market basket analysis dengan menggunakan algoritma apriori.

1.4 Manfaat Penelitian

Kesimpulan pada penelitian ini dapat diharapkan dapat bermanfaat untuk beberapa pihak, terutama bagi penulis, peneliti lain dan perusahaan Kompas.id.

1. Bagi penulis: Penulis dapat mengaplikasikan penerapan *Data Mining* berdasarkan *real case*, sehingga Penulis dapat berkontribusi untuk menyelesaikan masalah dan memberikan rekomendasi pemasaran yang efektif.
2. Bagi perusahaan: Sebagai rekomendasi pemasaran kepada Kompas.id untuk memberikan rekomendasi pemasaran berdasarkan data sehingga eksekusi penyelesaian masalah dapat lebih efektif.

3. Bagi Pihak Lain: Sebagai referensi pada penelitian dan bidang yang serupa serta dapat memberikan metode yang lebih akurat.

1.5 Batasan Masalah

Agar penelitian ini menjadi jelas dan terarah, batasan masalah pada penelitian ini dibatasi sebagai berikut:

1. Objek penelitian yang akan diteliti adalah artikel/berita yang dibaca oleh *register user* di Kompas.id.
2. Data yang digunakan merupakan database transaksi bacaan konsumen antara bulan Januari-September 2022.
3. Metode yang digunakan pada penelitian ini adalah algoritma Apriori dengan *Association Rule* dan pemberian rekomendasi untuk strategi pemasaran artikel/berita berdasarkan nilai *confidence*, *support*, *lift ratio*.

1.6 Definisi Istilah

1. *Artificial Intelligence* (AI): metode yang digunakan untuk meniru kecerdasan yang ada pada benda hidup maupun mati dalam menyelesaikan suatu masalah (Ahmad, 2017).
2. *Knowledge Discovery in Database*: serangkaian analisis terstruktur yang bertujuan untuk mendapatkan informasi yang akurat, baru, dan berguna dari data yang kompleks dan besar serta menemukan pola-pola yang terdapat di dalamnya (Zanuardi & Suprayitno, 2018).

3. *Clustering*: metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya (Suyanto, 2019).
4. *Klasifikasi*: men-generalisasi struktur yang telah didapatkan untuk diimplementasikan pada data baru. (Suyanto, 2019).
5. *Regresi*: menemukan suatu fungsi yang dibuat model data dengan kesalahan prediksi sesedikit mungkin (Suyanto, 2019).
6. *Machine learning*: cabang dari AI/kecerdasan buatan yang memberikan computer kemampuan untuk belajar tanpa secara eksplisit diprogram (Bickel dkk., 2023).
7. *Deteksi Anomali*: mengidentifikasi data yang tidak umum seperti terdapat *outlier*, perubahan atau deviasi yang mungkin esensial dan perlu diidentifikasi lagi lebih detail (Suyanto, 2019).
8. *Association rule learning*: pembelajaran aturan asosiasi yang berfungsi untuk mencari hubungan antar variable (Suyanto, 2019).
9. *Summarization*: memberi gambaran data lebih sederhana yang dapat berupa visualisasi atau *report* (Suyanto, 2019).
10. *Algorithm*: suatu prosedur yang digunakan untuk menyelesaikan masalah atau melakukan komputasi. Algoritma berfungsi sebagai daftar instruksi yang tepat yang melakukan tindakan tertentu secara bertahap dalam rutinitas berbasis perangkat keras atau lunak (Gillis, 2022).
11. *Preprocessing data*: cara memperbaiki data supaya lebih tertata, akurat dan bersih sebelum diproses (Shevira dkk., 2022).

BAB II

KAJIAN TEORI

2.1 *Data Mining*

2.1.1 *Pengertian Data Mining*

Data Mining merupakan sebuah kelas teknik analitis yang melampaui statistik dan bertujuan untuk memeriksa sejumlah data yang besar. Teknik *data mining* digunakan untuk menganalisis dan mengekstraksi data penting dari *database* untuk mendapatkan informasi baru dan berharga dalam bentuk pola dan aturan dari hubungan antar elemen data (Hirji, 2001). Metode yang digunakan pada *data mining* adalah irisan dari AI, statistika, *machine learning*, dan *database systems* (Kesavaraj & Sukumaran, 2013).

Beberapa tahun terakhir data semakin memiliki banyak variasi dan kompleks dengan volume yang meningkat. Pada tahun 2023 terdapat 120 zettabytes, sedangkan pada 2024 diperkirakan terdapat 147 zettabytes, kemudian pada 2025 diperkirakan meningkat sebesar 181 zettabytes, peningkatan volume data ini cukup pesat secara eksponensial (Poore dkk., 2022). Berdasarkan informasi tersebut, saat ini terdapat istilah *Big Data*. Big data memiliki 4 karakteristik yaitu *high volume, high variety, high velocity, dan high veracity*. Namun, *big data* dapat dianalisis sehingga perusahaan dapat mengambil insight untuk strategi bisnis yang menguntungkan dan lebih baik. Akan tetapi, berdasarkan *big data* tersebut bagaimana cara menemukan pola dan relasi data secara cepat?

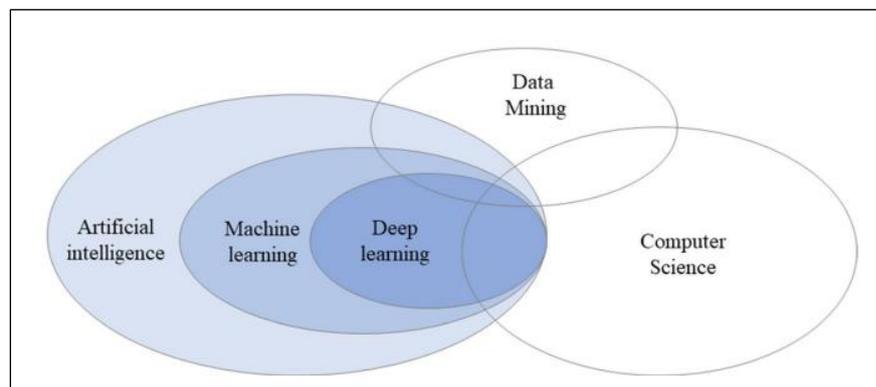
Data mining adalah sebuah proses yang relatif cepat dan mudah untuk mengekstraksi informasi dari *dataset* dan ditransformasikan ke struktur yang dapat

dipahami untuk menemukan *insight*, pola dan atau relasi secara otomatis (Suyanto, 2019). Proses ini merupakan proses komputasi dari mengamati pola pada data berukuran besar. Tugas *data mining* yaitu secara otomatis atau semi otomatis menganalisis besar jumlah data untuk mengekstrak pola menarik yang sebelumnya tidak diketahui (Kesavaraj & Sukumaran, 2013).

Implementasi *data mining* banyak digunakan pada berbagai jenis bidang seperti seni, sains, jurnalistik, politik, kesehatan, *marketing*, bisnis, dan lain-lain. Pada bidang *marketing* dan bisnis untuk mendapatkan profit yang maksimal dibutuhkan biaya seminimal mungkin. Contohnya seperti dalam pemilihan vendor, pemilihan model pemasaran yang efektif dan tepat sasaran, serta deteksi kecurangan yang mungkin dilakukan oleh pelanggan. Penerapan *data mining* dalam bidang *marketing* dan bisnis dapat menggunakan *market basket analysis* atau dikenal sebagai *association rule* yang berfungsi untuk menemukan keterkaitan antar data. Sehingga dapat digunakan untuk menempatkan produk yang memiliki asosiasi yang kuat di tempat yang berdekatan sehingga pelanggan dapat menemukannya lebih mudah. Selain itu ada *churn prediction* yang fungsinya adalah untuk memprediksi pelanggan yang loyal dan tidak loyal agar bisa menentukan model promosi seperti apa yang paling sesuai untuk tiap kategori pelanggan. Pada bidang seni dan hiburan, *data mining* juga dapat membuat sistem klasifikasi otomatis dengan mengelompokkan jutaan lagu dalam sejumlah *genre*. Sehingga para pendengar tidak perlu memakan waktu untuk menentukan *genre* lagu yang digemari. Pada bidang sains dan teknik, *data mining* juga dapat menyelesaikan permasalahan seperti genetika, medis, teknik elektro, dan lain-lain. Misalnya pada bidang biologi pencocokan Fotonan genetika atau yang biasa disebut teknik

sequence mining dapat dilakukan oleh *data mining* untuk melakukan pemetaan hubungan darah antar manusia. Metode yang biasanya digunakan adalah *multifactor dimensionality reduction* (Suyanto, 2019).

Berdasarkan beberapa literatur, para ahli mendefinisikan bahwa *data mining* menggunakan berbagai macam teknik untuk menemukan pola-pola penting dari *big data*. Sehingga, teknik *machine learning* atau pembelajaran mesin merupakan teknik yang paling sering digunakan dalam penggalian data atau *data mining*. Akan tetapi, *machine learning* juga berdampingan dengan *deep learning* dan *AI*. Berikut merupakan gambaran yang mendefinisikan *data mining*, *machine learning*, *AI*, dan *deep learning* pada Gambar 2.1.



Gambar 2.1 Gambaran AI, ML, DL, *Data Mining*, dan *Computer Science*

AI adalah suatu ilmu seperti biologi atau matematika yang mempelajari cara membangun suatu program cerdas yang dapat secara kreatif menyelesaikan masalah dengan meniru kemampuan manusia. Sedangkan *machine learning* adalah *subset* dari AI yang memberikan sistem kemampuan untuk belajar dan meningkatkan diri dari pengalaman tanpa perlu diprogram secara eksplisit. Sedangkan *deep learning* atau adalah subbagian dari *machine learning* (W. Zhang dkk., 2021).

2.1.2 Pengelompokan *Data Mining*

Data Mining dibagi menjadi beberapa bagian sesuai dengan tugas yang dikerjakan, di antaranya:

1. Deskripsi

Tujuan dari deskripsi adalah untuk menemukan kecenderungan suatu data, peneliti dapat mendeskripsikan pola dalam data.

2. Estimasi

Pada estimasi, variabel target estimasi lebih memiliki tipe data numerik daripada kategori. Model dibuat dengan baris data atau record yang menampung nilai dari variabel target sebagai nilai prediksi. Kemudian, estimasi nilai dari variabel target tersebut dijadikan nilai variabel prediksi.

3. Prediksi

Prediksi adalah nilai yang dihasilkan pada masa mendatang. Pada prediksi dapat juga digunakan metode dan teknik pada kelompok klasifikasi dan estimasi.

4. Klasifikasi

Klasifikasi memiliki target variabel kategori. Misalnya, penggolongan umur dapat dibagi menjadi tiga kategori, contoh umur muda, umur dewasa, dan umur tua.

5. Pengklasteran

Pengklasteran adalah pengelompokan record, observasi, dan membentuk kelas objek-objek yang serupa. Klaster merupakan gabungan record yang terdapat keserupaan dan mempunyai ketidakserupaan record

pada kluster yang lain. Terdapat perbedaan antara pengklasteran dan klasifikasi, di mana pada klasifikasi tidak memerlukan variabel target.

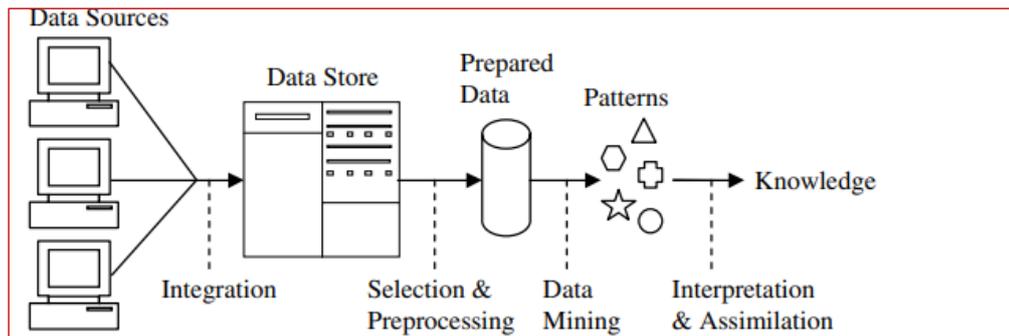
Pada klasifikasi tidak terdapat variabel target. Sedangkan pada pengklasteran juga tidak dilakukan klasifikasi, estimasi atau prediksi nilai berdasarkan variabel target, namun algoritma pengklasteran melakukan pembagian pada semua data menjadi kelompok yang terdapat keserupaan atau homogen, pada kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan pada record kelompok lain bernilai minimal.

6. Asosiasi

Asosiasi bertugas sebagai penemu atribut yang ada dalam satu waktu. Pengaplikasian dari asosiasi salah satunya adalah *association rule* sebagaimana yang akan dibahas pada penelitian ini. *Association rule* meliputi dua tahap yaitu menemukan kombinasi berdasarkan item yang paling sering terjadi pada suatu itemset dan mendefinisikan condition dan result atau conditional *association rule*. Pada *association rule* terdapat ukuran kepercayaan yang didapatkan berdasarkan dari hasil processing data dengan rumus tertentu (Rerung, 2018).

2.1.3 Tahapan *Data Mining*

Pada *data mining* akan dihasilkan ekstraksi informasi atau pola yang menarik berdasarkan data dalam *database*. Istilah lain yang biasa digunakan untuk *data mining* adalah *KDD*. Penjelasan mengenai *KDD* dapat dilihat pada Gambar 2.2 (Rerung, 2018).



Gambar 2.2 Knowledge Discovery in Database (KDD)

1. Data Selection

Pada proses ini dilakukan seleksi terhadap kumpulan data operasional sebelum dilakukan pencarian informasi dalam *KDD* dimulai. Dari data yang telah diseleksi tersebut akan dimanfaatkan untuk proses *data mining*, dan disimpan pada tempat terpisah dari basis data operasional (Rerung, 2018).

2. Pre-processing/Cleaning

Proses *cleaning* harus dikerjakan pada data yang dilakukan KDD sebelum melakukan *data mining*. Langkah yang dikerjakan ketika melakukan proses *data mining* di antaranya menghapus duplikasi data, cek data yang memiliki keinkonsistenan, dan membenahi kesalahan pengetikan pada data, proses *enrichment* atau proses “memperkaya” data dengan data yang sudah ada yang relevan (Gonrunescu, 2011).

3. Transformation

Coding merupakan prosedur transformasi data yang telah terseleksi, sehingga data itu telah relevan untuk dilakukan proses *data mining*. Proses *coding* pada *KDD* adalah proses inovatif dan tergantung pada tipe atau pola informasi yang akan digali pada basis data (Rerung, 2018).

4. *Data Mining*

Proses penggalian informasi atau pola menarik pada data yang telah dipilah berdasarkan metode atau teknik tertentu merupakan pengertian dari *data mining*. Terdapat banyak variasi teknik, algoritma, dan metode pada *data mining*. Proses *KDD* dan tujuan sangat terikat terhadap penentuan metode atau algoritma yang sesuai (Fayyad dkk., 1996).

5. Interpretation/Evaluation

Proses *data mining* yang menghasilkan pola informasi harus divisualisasikan dalam gambaran yang mudah dipahami oleh pihak yang bersangkutan. Interpretasi ini adalah langkah dari proses *KDD*. Pada tahap ini mengandung pengamatan apakah informasi atau pola yang didapatkan kontradiktif dengan hipotesis atau fakta yang sebelumnya ditemukan (Bramer, 2007).

2.2 *Association Rule*

2.2.1 *Pengertian Association Rule*

Association rule atau yang biasa disebut *market basket analysis* adalah salah satu teknik dalam *data mining* yang berfungsi untuk menemukan asosiasi atau keterkaitan data. *Association rule mining* adalah salah satu dari algoritma yang paling efektif untuk memecahkan berbagai macam masalah 2.. Teknik *association rule* ditemukan pada berbagai bidang seperti *telecommunications networks*, marketing, dan lain-lain (Son dkk, 2018).

Teknik ini dapat digunakan untuk menempatkan barang-barang yang memiliki asosiasi kuat pada tempat yang berdekatan sehingga pelanggan dapat

dengan mudah menemukan barang yang diperlukan. Misalnya, Polhuk, Opini, dan Humaniora diletakkan di tampilan yang berdekatan karena asosiasi dari ketiga barang tersebut sangat kuat (Suyanto, 2019). *Association rule* mining dapat didefinisikan sebagai $I = \{i_1, i_2, \dots, i_m\}$ di mana, I adalah himpunan dari beberapa *item*. Misalnya, produk seperti Foto, Opini, Internasional untuk dibeli dari toko dan $A_i = v$ adalah sebuah *item*. Di mana v adalah nilai domain dari atribut A_i , dalam relasi $R(A_1, \dots, A_n)$ (Rerung, 2018)

X merupakan sebuah *itemset* apabila X merupakan himpunan bagian dari I. Sebagai contoh, sebuah himpunan barang untuk dibeli dari toko adalah sebuah himpunan *item*, dan himpunan dari $A_i = v$ adalah sebuah himpunan *item* untuk relasi $R(PID, A_1, A_2, \dots, A_n)$, PID atau *primary key identifier* digunakan untuk mengidentifikasi pengenalan unik untuk produk. $D = \{t_1, t_{i+1}, \dots, t_n\}$ adalah himpunan transaksi, yang disebut *database* transaksi, di mana tiap transaksi memiliki TID atau *transaction ID* adalah pengenalan unik untuk transaksi sebuah *t*-himpunan set $t = (tid, t - \text{himpunan set})$. Contohnya, terdapat transaksi dari seorang pelanggan: *tuple* (v_1, \dots, v_n) dari relasi $R(A_1, \dots, A_n)$ adalah sebuah transaksi. Sebuah transaksi t mengandung himpunan set X jika untuk seluruh *item*, di mana $i \in X$, i adalah *item* set- t . sebagai contoh, sebuah keranjang belanja mengandung seluruh *item* di X ketika akan dilakukan pembayaran dan untuk tiap $A_i = v_i$ di X, v_i muncul pada posisi i pada *tuple* (v_1, \dots, v_n) (Rerung, 2018)

Terdapat 2 *natural lattice* struktur himpunan item pada 2^I , yaitu struktur *subset* atau *superset*. Contohnya, *item* dari sesuatu jenis tertentu seperti pakaian, peralatan keras, perabotan, dan lain-lain serta dari pakaian, pakaian anak-anak, pakaian wanita, pakaian pria, pakaian balita, dan lain-lain. *Lattice* ini membantu

dalam mengidentifikasi pola-pola atau hubungan antara beberapa *item* yang ada dalam transaksi dan dapat digunakan untuk tujuan analisis atau pengambilan keputusan dalam bisnis atau domain lainnya. Dengan kata lain *association rule mining* dapat membantu kita memahami pola pembelian atau preferensi pelanggan, rekomendasi produk atau penempatan barang dalam toko, dan sebagainya (C. Zhang & Zhang, 2002).

2.2.2 Ukuran Kepercayaan *Rule*

Terdapat dua ukuran kepercayaan *rule* atau *interestingness measure* yang memberikan informasi kepastian pada tingkat fungsi pada *rule* yang ditemukan, (Bramer, 2017) yaitu:

1. *Support*

Support adalah minimal presentase dari transaksi dalam *database* yang mengandung X dan Y yaitu $X \cup Y$. *Support* juga merupakan suatu parameter yang menunjukkan ukuran dominasi suatu *item* atau *itemset* dari seluruh transaksi yang ada (Rerung, 2018). Misalnya, terdapat himpunan item X pada *database* transaksi D yang memiliki *support*, disimbolkan sebagai $supp(X)$. Transaksi di D berisi X, atau dapat dituliskan sebagai berikut:

$$Supp(X) = \frac{|X(t)|}{|D|} \quad (2.1)$$

di mana $X(t) = \{t \text{ di } D | t \text{ mengandung } X\}$.

Himpunan *item* dalam *database* D disebut sebagai himpunan item yang besar atau sering muncul jika *support*-nya sama dengan atau lebih besar dari ambang batas *minimal support* (*minimum support*) yang diberikan

oleh pengguna. Negasi dari himpunan item X adalah $\neg X$. Nilai *support* dari $\neg X$ adalah $supp(\neg X) = 1 - supp(X)$. *Association rule* merupakan implikasi dari $X \rightarrow Y$, dimana himpunan item X dan Y tidak beririsan. Tiap *association rule* memiliki 2 ukuran yaitu *support* dan *confidence*, didefinisikan dengan *support* pada *rule* $X \rightarrow Y$ adalah *support* dari $X \cup Y$ (C. Zhang & Zhang, 2002).

2. *Confidence*

Confidence adalah minimum persentase dari transaksi $X \cap Y$. *Confidence* juga disebut sebagai parameter yang menunjukkan relasi antar item secara kondisional (contohnya apakah item Y memiliki frekuensi penjualan yang tinggi jika orang *membeli* item X). Untuk mendapatkan *association rule* maka harus didapatkan nilai *support* yang telah ditentukan. *Support* Untuk mendapatkan nilai *support* dapat dicari seluruh aturan yang jumlah *support* lebih besar sama dengan *minimum support*. Cara ini juga dapat dimanfaatkan untuk mendapatkan nilai *confidence*. Nilai *confidence* didapatkan berdasarkan nilai *support* aturan pada transaksi yang ada. Nilai *confidence* didefinisikan dengan: *confidence* dari *rule* $X \rightarrow Y$ merupakan nilai $conf(X \rightarrow Y)$ sebagai rasio $|X \cup Y|(t)/|X(t)|$, atau $support(X \cup Y)/support(X)$. Kesimpulannya, *confidence* merupakan seberapa kuat nilai implikasi (C. Zhang & Zhang, 2002).

Menurut Agrawal & Srikant (1994), *framework* dari *support-confidence* ditunjukkan sebagai berikut: misalkan I adalah sebuah himpunan item dalam sebuah *database* D , $X, Y \subseteq I$ menjadi himpunan item, $X \cap Y = \emptyset$, $supp(X) \neq 0$. *Minimum support (minimum support)* dan *minimal*

confidence (minconf) diberikan berdasarkan pengguna/ahli. Maka $X \rightarrow Y$ merupakan *rule* yang *valid* apabila:

$$\text{support}(X \cup Y) \geq \text{minsupp} \quad (2.2)$$

dan

$$\text{conf} = \frac{\text{support}(X \cup Y)}{\text{support}(X)} \geq \text{minconf} \quad (2.3)$$

Di mana $\text{conf}(X \rightarrow Y)$ merujuk pada *confidence rule* dari $X \rightarrow Y$ berdasarkan persamaan 2.3. *Rule* dari *association rule mining* dapat dibagi menjadi dua sub:

- a. Menghasilkan semua himpunan item yang memiliki *support* yang lebih besar atau sama dengan *minimum support* yang ditentukan oleh pengguna. Yaitu menghasilkan semua himpunan item yang sering muncul.
- b. Menghasilkan semua aturan yang memiliki minimum *confidence* dengan cara sederhana berikut: untuk setiap himpunan item yang sering muncul X , dan $B \subset X$, misalkan $A = X - B$. Jika *rule* dari *confidence* $A \rightarrow B$ lebih besar dari atau sama dengan *minimum confidence*, maka *rule* tersebut dapat dinyatakan sebagai *rule* yang *valid*. Penggunaan pada *support-confidence framework* dicontohkan sebagai berikut.

Tabel 2.1 Transaksi Database

TID	Items
200	Opini, Polhuk, Foto
300	Internasional, Polhuk, Ekonomi
400	Opini, Internasional, Polhuk, Ekonomi
500	Internasional, Ekonomi

Misalkan himpunan semesta item $I = \{\text{Opini, Internasional, Polhuk, Foto, Ekonomi}\}$ dan semesta transaksi $TID = \{200, 300, 400, 500\}$. Pada Tabel 2.1, nilai 200, 300, 400, 500 adalah pengenal unik dari 4 transaksi. Setiap baris dalam tabel dapat dianggap sebagai satu transaksi. Dari transaksi tersebut kita dapat mengidentifikasi *association rule* dari transaksi ini menggunakan *framework support-confidence*. Misalkan minimal *support* = 50% (untuk menjadi himpunan item yang sering muncul, sebuah himpunan item harus muncul dalam setidaknya 2 transaksi) dan *minimum confidence* = 60% (untuk menjadi aturan yang memiliki *confidence* tinggi atau *valid*) (C. Zhang & Zhang, 2002). Dengan menggunakan *framework support-confidence*, dapat dilakukan *association rule mining* dengan dua tahap sebagai berikut:

- i. Hitung frekuensi k-himpunan item. Pada Tabel 2.1, item {Opini} muncul dalam dua transaksi, TID = 200 dan TID = 400. Frekuensinya ada 2, dan *support*-nya, $support(Opini)$ adalah 50% yang sama dengan *minimum support* = 50%. Item {Internasional} muncul dalam 3 transaksi, TID = 300, TID = 400, dan TID = 500. Frekuensi-nya ada 3, dan *support* dari $support(Polhuk) = 75\%$ yang mana lebih besar dari *minimum support*. Item {Polhuk} muncul dalam 3 transaksi, TID = 200, TID = 300, dan TID = 400. Frekuensi-nya adalah 3, dan *support* dari $supp(Polhuk)$ adalah 75% lebih besar dari *minimum support*. Item {Foto} muncul pada 1 transaksi, TID =

200. Frekuensi-nya adalah 1. Kemudian. Dan *support*-nya, *supp(Ekonomi)* muncul dalam 3 transaksi, TID = 300, TID = 400, dan TID = 500. Frekuensi-nya adalah 3 dan *support*-nya atau *support(Ekonomi)* adalah 75% yang lebih besar dari *minimum support*. Hal ini dapat disimpulkan dalam Tabel 2.2.

Tabel 2.2 Himpunan Item-1 Dalam *Database*

Item	Frekuensi	> <i>minsupport</i>
Opini	2	Y
Internasional	3	Y
Polhuk	3	Y
Foto	1	N
Ekonomi	3	Y

Pada Tabel 2.2, himpunan item {Opini, Internasional} muncul dalam 1 transaksi, TID = 400. Frekuensi-nya adalah 1, dan *support*-nya, *support (Opini \cup Internasional)*, adalah 25% yang kurang dari *minimum support* = 50%. Berdasarkan rumus, *Opini \cup Internasional* merupakan {Opini, Internasional}. Himpunan item {Opini, Polhuk} muncul dalam 2 transaksi, TID = 200 dan TID = 400, frekuensi-nya sebesar 2, dan *support*-nya, *support(Opini \cup Polhuk)* adalah 50 % yang sama dengan *minimum support* = 50%. Himpunan set {Opini, Foto} muncul dalam 1 transaksi pada TID = 200. Frekuensi-nya adalah 1 dan *support*-nya atau *support(Opini \cup Ekonomi)* muncul dalam

1 transaksi pada TID = 400. Frekuensi-nya adalah 1, dan *support*-nya atau $support(Opini \cup Ekonomi)$ adalah 25% yang kurang dari *minimum support* = 50%. Himpunan item {Internasional, Polhuk} muncul dalam 2 transaksi, TID =300 dan TID = 400. Frekuensi-nya adalah 2 dan $support(Internasional \cup Polhuk)$ adalah 50% yang sama dengan *min support*. Perhitungan ini dapat ditunjukkan pada Tabel 2.3.

Tabel 2.3 Himpunan Set-2 Dalam Database

Himpunan item	Frekuensi	> <i>minsupport</i>
{Opini, Internasional}	1	N
{Opini, Polhuk}	2	Y
{Opini, Foto}	1	N
{Opini, Ekonomi}	1	N
{Internasional, Polhuk}	2	Y
{Internasional, Ekonomi}	3	Y
{Polhuk, Foto}	1	N
{Polhuk, Ekonomi}	2	Y

Dengan cara yang sama, himpunan set-3 dan himpunan set-4 dapat diperoleh. Hal ini dituliskan dalam Tabel 2.4 dan Tabel 2.5.

Tabel 2.4 Himpunan Set-3 Dalam *Database*

Himpunan set	Frekuensi	> <i>minsupport</i>
{Opini, Internasional, Polhuk}	1	N
{Opini, Internasional, Ekonomi}	1	N
{Opini, Polhuk, Foto}	1	N
{Opini, Polhuk, Ekonomi}	1	N
{Internasional, Polhuk, Ekonomi}	2	Y

Tabel 2.5 Himpunan Item-4 Dalam *Database*

Himpunan Item	Frekuensi	> <i>minsupp</i>
{Opini, Internasional, Polhuk, Ekonomi}	1	N

Artinya, himpunan set-5 dalam *database* yaitu *null*. Berdasarkan definisi di atas, {Opini}, {Internasional}, {Polhuk}, {Ekonomi}, {Opini, Polhuk}, {Internasional, Polhuk}, {Internasional, Ekonomi}, {Polhuk, Ekonomi}, {Internasional, Polhuk, Ekonomi} dalam Tabel 2.1 merupakan himpunan item yang himpunan item yang sering muncul atau *frequent itemset* (C. Zhang dkk., 2019).

- ii. Menghasilkan semua *association rule* dari *frequent itemset*. Karena tidak terdapat *frequent itemset* pada Tabel 2.5, himpunan *itemset* ke-4 tidak memberikan *association rule* yang *valid*. Di Tabel 2.4, terdapat satu *frequent itemset*,

{Internasional, Polhuk, Ekonomi} dengan $support (Internasional \cup Polhuk \cup Ekonomi) = 50\% = minimum support$. Untuk *frequent itemset* {Internasional, Polhuk, Ekonomi}, karena $support (Internasional \cup Polhuk \cup Ekonomi) = \frac{2}{2} = 100\%$ lebih besar dari $minimum confidence = 60\%$, di mana $Internasional \cup Polhuk \rightarrow Ekonomi$ sehingga dapat diekstrak menjadi *rules* yang valid.

Dengan cara yang sama, karena $\frac{Support(Internasional \cup Polhuk \cup Ekonomi)}{Support(Internasional \cup Ekonomi)} = \frac{2}{3} = 66.6\%$ lebih besar dari $minimum confidence$, $Polhuk \cup Ekonomi \rightarrow Internasional$ dapat diekstrak sebagai *rule* yang valid. Karena $\frac{support(Internasional \cup Polhuk \cup Ekonomi)}{support(Internasional)} = \frac{2}{3} = 66.7\%$ lebih besar dari $minimum confidence$, $Internasional \rightarrow Polhuk \cup Ekonomi$ dapat diekstrak sebagai *rule* yang dihasilkan dari (Internasional, Polhuk, Ekonomi) terdaftar dalam Tabel 2.6 dan Tabel 2.7 (C. Zhang & Zhang, 2002).

Tabel 2.6 Association Rule Dengan 1-Item Konsekuensi 3 Himpunan Set

No.	Rule	Confidence	Support	> minconfidence
Rule 1	$Internasional \cup Polhuk \rightarrow Ekonomi$	100%	50%	Y
Rule 4	$Internasional \cup Ekonomi \rightarrow Polhuk$	66.7%	50%	Y
Rule 5	$Polhuk \cup Ekonomi \rightarrow Internasional$	100%	50%	Y

Tabel 2.7 Association Rules dengan 2-Item Konsekuensi 3 Himpunan Item

No.	Rule	Confidence	Support	> minconfidence
Rule 3	<i>Internasional</i> → <i>Polhuk</i> ∪ <i>Ekonomi</i>	66.7%	50%	Y
Rule 4	<i>Polhuk</i> → <i>Internasional</i> ∪ <i>Ekonomi</i>	66.7%	50%	Y
Rule 5	<i>Ekonomi</i> → <i>Internasional</i> ∪ <i>Ekonomi</i>	66.7%	50%	Y

Selain itu, kita dapat menghasilkan semua *association rule* dari *frequent itemset* ke-2 di Tabel 2.3. Hal ini dituliskan pada Tabel 2.8, Tabel 2.9. Berdasarkan Tabel 2.9, ditunjukkan bahwa ketika terjadi transaksi atau kejadian yang melibatkan {Polhuk}, sekitar 66,7% dari transaksi tersebut juga melibatkan {Internasional}. Dengan tingkat *support* sebesar 50%, aturan ini cukup kuat dan dapat diandalkan. Keberhasilan prediksi atau asosiasi dari {Polhuk} ke {Internasional} memiliki tingkat *confidence* sebesar 66,7%.

Tabel 2.8 Association Rule Untuk {Opini, Polhuk}

No.	Rule	Confidence	Support	> minconfidence
Rule 7	<i>Opini</i> → <i>Polhuk</i>	100%	50%	Y
Rule 8	<i>Polhuk</i> → <i>Opini</i>	66.7%	50%	Y

Berdasarkan definisi di atas, daftar 14 *association rules* di atas dapat diekstraksi sebagai *rules* yang *valid* untuk Tabel 2.1

Tabel 2.9 Association Rules Untuk {Internasional, Polhuk}

No.	Rule	Confidence	Support	> minconfidence
Rule 9	<i>Internasional</i> → <i>Polhuk</i>	66.7%	50%	Y
Rule 10	<i>Polhuk</i> → <i>Internasional</i>	66.7%	50%	Y

Berdasarkan Tabel 2.9, ditunjukkan bahwa ketika terjadi transaksi atau kejadian yang melibatkan {Polhuk}, sekitar 66,7% dari transaksi tersebut juga melibatkan {Internasional}. Dengan tingkat *support* sebesar 50%, aturan ini cukup kuat dan dapat diandalkan. Keberhasilan prediksi atau asosiasi dari {Polhuk} ke {Internasional} memiliki tingkat *confidence* sebesar 66,7%.

Tabel 2.10 Association Rule Untuk {Internasional, Ekonomi}

No.	Rule	Confidence	Support	> minconfidence
Rule 11	<i>Internasional</i> → <i>Ekonomi</i>	100%	75%	Y
Rule 12	<i>Ekonomi</i> → <i>Polhuk</i>	100%	75%	Y

Berdasarkan Tabel 2.10, ditunjukkan bahwa setiap terjadi transaksi yang melibatkan berita {Ekonomi}, transaksi tersebut memiliki *confidence* sebesar 100% juga melibatkan {Polhuk}/ tingkat *support* sebesar 75% menunjukkan bahwa aturan ini terjadi dalam 75% dari seluruh transaksi yang melibatkan

{Ekonomi}. Aturan ini juga sangat kuat atau dengan tingkat *confidence* sebesar 100% yang menandakan dapat diandalkan.

Tabel 2.11 *Association Rules* Untuk {Polhuk, Ekonomi}

No.	Rule	Confidence	Support	> minconfidence
Rule 13	<i>Polhuk</i> → <i>Ekonomi</i>	66.7%	50%	Y
Rule 14	<i>Ekonomi</i> → <i>Polhuk</i>	66.7%	50%	Y

Berdasarkan tabel 2.11, ditunjukkan bahwa terjadi transaksi yang melibatkan {Ekonomi} sebesar 66,7% dari transaksi tersebut juga melibatkan {Polhuk}. Dengan tingkat *support* sebesar 50%, aturan ini cukup kuat dan dapat diandalkan setidaknya dalam setengah dari transaksi yang melibatkan berita {Ekonomi}. Keberhasilan prediksi atau asosiasi dari {Ekonomi} ke {Polhuk} memiliki tingkat *confidence* sebesar 66,7%.

3. *Lift Ratio*

Lift ratio merupakan ukuran untuk mendapatkan *association rule* yang telah ditemukan. Sebagai acuan apakah *association rules valid* atau tidak dapat menggunakan ukuran *lift ratio* (Fauzy dkk, 2016). Ukuran ini digunakan untuk mengukur seberapa besar kemungkinan kemunculan item Y jika item X sudah muncul, dibandingkan dengan kemunculan item Y secara acak. *Lift ratio* dihitung dengan membagi probabilitas bersama item X dan Y muncul $P(X \cap Y)$ dengan probabilitas item X muncul dikalikan dengan probabilitas item Y muncul secara terpisah $(X) \times P(Y)$. Semakin besar

nilai *lift ratio* maka semakin besar pula hubungan antara item A dan item B (Agrawal, 1994). Rumus *lift ratio* dituliskan sebagai berikut:

$$Lift\ Ratio = \frac{Confidence(X,Y)}{Support(X) \times Support(Y)} \quad (2.4)$$

2.3 Algoritma Apriori

Algoritma apriori pertama kali dikembangkan oleh Agrawal dan Srikan pada tahun 1994. Algoritma ini adalah salah satu cara inovatif untuk menemukan *association rules* pada skala besar yang memungkinkan hasil implikasi yang terdiri dari lebih dari satu item (Husna dkk., 2019). *Association rule mining* adalah teknik yang digunakan untuk menemukan hubungan antara sekumpulan *variable* yang sangat besar dalam basis data. Maka dari itu, dibutuhkan algoritma yang cepat untuk menyelesaikan tugas ini (Agrawal & Srikant, 1994).

Konsep dari algoritma apriori digunakan untuk mengekstraksi *frequent itemset* atau himpunan set dari basis data besar, kemudian untuk menghasilkan aturan asosiasi dari setiap *frequent itemset*. Informasi tersebut dapat digunakan sebagai basis keputusan terhadap aktivitas *marketing* seperti harga promosi, produk *bundling*, atau penempatan produk (C. Zhang & Zhang, 2002). Algoritma apriori melakukan beberapa kali proses pada *database* yang diberikan.

Himpunan *frontier* untuk setiap proses terdiri dari himpunan item yang diperpanjang selama proses tersebut. *Frontier* adalah himpunan item yang diperluas selama tiap iterasi algoritma. Dalam setiap iterasi, algoritma mencari kandidat himpunan item baru berdasarkan himpunan item yang ditemukan pada iterasi sebelumnya dan memeriksa apakah kandidat himpunan item tersebut sering muncul dalam *database*. Kandidat himpunan set yang memenuhi batas *minimum support*

akan ditambahkan ke himpunan *frontier* untuk iterasi berikutnya, sedangkan yang tidak, akan dihapus dari *frontier set*.

Setiap himpunan item dihubungkan dengan sebuah penghitung yang menyimpan jumlah transaksi di mana himpunan set yang sesuai muncul. Penghitung ini diinisialisasi menjadi nol saat himpunan set dibuat. Pada awalnya, *frontier set* terdiri dari satu elemen kosong. Kemudian di akhir, *support* untuk kandidat himpunan item dibandingkan dengan *minimum support* untuk menentukan apakah itu adalah *frequent* himpunan item. Pada saat yang sama, ditentukan apakah itemset ini harus ditambahkan ke *frontier set* untuk proses berikutnya. Algoritma berakhir ketika *frontier set* menjadi kosong. Jumlah *support count* untuk himpunan itemset tetap saat himpunan set ditambahkan ke *frequent/frontier set*.

Akan tetapi pada *database* besar yang diberikan, penggunaan algoritma apriori untuk mengidentifikasi *frequent itemsets* melibatkan pencarian dengan sedikit informasi dalam ruang dengan jumlah item dan himpunan item yang mungkin eksponensial. Algoritma ini dapat mengalami *overhead* komputasi yang besar jika jumlah *frequent itemsets* sangat besar. Sebagai contoh, misalnya ada 1000 item pada *database* besar yang diberikan, dengan rata-rata jumlah item pada tiap transaksi adalah 6. Maka hampir 10^{15} kemungkinan himpunan item yang harus dihitung pada *database* tersebut (C. Zhang & Zhang, 2002).

Penjelasan mengenai algoritma apriori dijelaskan sebagai berikut pada data Tabel 2.1, dengan *minimum support* = 50%. Pertama, himpunan item-1 {Opini}, {Internasional}, {Polhuk}, {Foto}, dan {Ekonomi}. Dihasilkan sebagai kandidat pada proses pertama pada *dataset*, di mana *count* Opini = 2, *count* Internasional = 3, *count* Polhuk = 3, *count* Foto = 1 dan *count* Ekonomi = 3. Karena *minimum*

$support = 50\%$ dan $dbsize$ atau ukuran database = 4, {Opini}, {Internasional}, {Polhuk}, {Ekonomi} adalah *frequent itemset*. Ketiga himpunan set tersebut terdaftar pada daftar Tabel 2.12.

Tabel 2.12 Frekuensi Himpunan Set-1 Pada *Dataset* Tabel 2.1

Himpunan Item	Frekuensi	$> \text{minsupport}$
{Opini}	2	Y
{Internasional}	3	Y
{Polhuk}	3	Y
{Ekonomi}	3	Y

Kedua, frekuensi himpunan item-1 yaitu {Opini}, {Internasional}, {Polhuk}, {Ekonomi} ditambahkan ke frontier set F, dan proses kedua dimulai pada *dataset* untuk mencari kandidat himpunan set-2. Setiap kandidat seperti ini adalah subset dari F. kandidat himpunan set-2 adalah {Opini, Internasional}, {Opini, Polhuk}, {Opini, Ekonomi}, {Internasional, Polhuk}, {Internasional, Ekonomi} dan {Polhuk, Ekonomi}, di mana $Opini \cup count.Internasional = 1$, $Opini \cup count.Polhuk = 2$, $Opini \cup Polhuk$, $Internasional \cup Polhuk$, $Internasional \cup Polhuk$, $Polhuk \cup Ekonomi$ adalah *frequent itemset* yang terdaftar pada Tabel 2.13.

Tabel 2.13 Frekuensi Himpunan Item-2 Pada *Dataset* Di Tabel 2.1

Himpunan Item	Frekuensi	$> \text{minsupport}$
{Opini, Polhuk}	2	Y
{Internasional, Polhuk}	2	Y

Lanjutan Tabel 2.13 Frekuensi Himpunan Item-2 Pada *Dataset* Di Tabel 2.1

{ Internasional, Ekonomi }	3	Y
{ Internasional, Ekonomi }	2	Y

Ketiga, himpunan set frekuensi 1 dan 2 tersebut ditambahkan ke dalam *frontier* set F, dan tahap ketiga dimulai dengan mencari kandidat himpunan set-3 di seluruh *dataset*. Frequent himpunan set-3 tercantum pada Tabel 2.14.

Tabel 2.14 Frekuensi Himpunan Item-3 Pada *Dataset*

Himpunan Item	Frekuensi	> <i>minsupport</i>
{ Internasional, Polhuk, Ekonomi }	2	Y

Keempat, frekuensi himpunan set 1, 2, dan 3 ditambahkan ke dalam set *frontier* F, dan tahap keempat dimulai pada *dataset* untuk mencari kandidat himpunan set-4. Tidak ada frekuensi himpunan set-4 yang ditemukan, sehingga algoritma diakhiri (C. Zhang & Zhang, 2002).

2.4 Kajian Integrasi Topik dengan Al-Quran/Hadits

Allah SWT berfirman dalam surat Al-Hujurat ayat 12 (Kemenag RI, 2023):

يَا أَيُّهَا الَّذِينَ آمَنُوا اجْتَنِبُوا كَثِيرًا مِّنَ الظَّنِّ إِنَّ بَعْضَ الظَّنِّ إِثْمٌ وَلَا تَحْسَسُوا
وَلَا يَغْتَب بَّعْضُكُم بَعْضًا أَيُحِبُّ أَحَدُكُمْ أَنْ يَأْكُلَ لَحْمَ أَخِيهِ مَيْتًا فَكَرِهْتُمُوهُ وَاتَّقُوا
اللَّهَ إِنَّ اللَّهَ تَوَّابٌ رَّحِيمٌ

Artinya: “Wahai orang-orang yang beriman! Jauhilah banyak dari prasangka, sesungguhnya sebagian prasangka itu dosa dan janganlah kamu mencari-cari kesalahan orang lain dan janganlah ada di antara kamu yang menggunjing sebagian yang lain. Apakah ada di antara kamu yang suka memakan daging

saudaranya yang sudah mati? Tentu kamu merasa jijik. Dan bertakwalah kepada Allah, sesungguhnya Allah Maha Penerima tobat, Maha Penyayang.”

Tafsir mengenai pertanggungjawaban di akhirat bagi orang-orang yang menyebarkan berita palsu menyoroti aspek penting dalam ajaran agama Islam terkait dengan kejujuran, integritas, dan dampak sosial dari perbuatan yang merugikan tersebut (Kemenag RI, 2023). Pernyataan ini menegaskan bahwa tanggung jawab di hadapan Allah SWT di hari kiamat tidak hanya terbatas pada mereka yang secara aktif menciptakan berita palsu, melainkan juga melibatkan mereka yang turut serta menyebarkannya.

Dalam perspektif agama Islam, menyebarkan berita palsu atau dusta dapat dianggap sebagai perbuatan yang sangat merugikan dan bertentangan dengan nilai-nilai moral yang diajarkan dalam Al-Qur'an dan Hadis. Ajaran Islam menekankan pentingnya kejujuran, keadilan, dan kewaspadaan terhadap informasi yang tersebar. Menciptakan atau menyebarkan berita palsu dapat menimbulkan fitnah (penghinaan) terhadap individu atau kelompok, menyebabkan ketidakharmonisan sosial, dan merusak kepercayaan masyarakat.

Dalam konteks ini, tafsir tersebut memberikan gambaran bahwa Allah SWT akan menuntut pertanggungjawaban tidak hanya kepada para pencipta berita palsu, tetapi juga kepada penyebar-penyebarnya. Hal ini menunjukkan bahwa setiap individu memiliki tanggung jawab moral terhadap informasi yang mereka sebar, dan bahwa penyebaran berita palsu memiliki konsekuensi serius di akhirat.

Tafsir ini juga dapat dihubungkan dengan konsep amanah (tanggung jawab) dalam Islam, di mana setiap individu dianggap sebagai pemegang amanah untuk bertindak secara jujur dan adil. Oleh karena itu, menyebarkan berita palsu dianggap sebagai pengkhianatan terhadap amanah tersebut, dan hal ini akan menjadi bahan

pertimbangan ketika seseorang diminta pertanggungjawaban di hadapan Allah SWT.

Dengan demikian, tafsir ini mencerminkan nilai-nilai moral dan etika Islam yang mengajarkan kehati-hatian dalam menyikapi informasi, kejujuran dalam berkomunikasi, dan tanggung jawab terhadap dampak sosial dari perbuatan-perbuatan yang dapat merugikan orang lain.

Disebutkan dalam Al-Qur'an ciri-ciri orang munafik salah satunya adalah suka menipu. dijelaskan dalam surat Al-Baqarah ayat 8-9:

وَمِنَ النَّاسِ مَنْ يَقُولُ آمَنَّا بِاللَّهِ وَالْيَوْمِ الْآخِرِ وَمَا هُمْ بِمُؤْمِنِينَ ۝ ۸ يُخَادِعُونَ اللَّهَ
وَالَّذِينَ آمَنُوا ۚ وَمَا يُخَادِعُونَ إِلَّا أَنفُسَهُمْ وَمَا يَشْعُرُونَ ۝

Artinya: "Dan di antara manusia ada yang berkata, "Kami beriman kepada Allah dan hari akhir," padahal sesungguhnya mereka itu bukanlah orang-orang yang beriman. Mereka menipu Allah dan orang-orang yang beriman, padahal mereka hanyalah menipu diri sendiri tanpa mereka sadari,"

Dijelaskan dalam surat At-Taubah ayat 68 bahwa munafik adalah akhlak tercela:

وَعَدَّ اللَّهُ الْمُنَافِقِينَ وَالْمُنَافِقَاتِ وَالْكُفَّارَ نَارَ جَهَنَّمَ خَالِدِينَ فِيهَا ۚ هِيَ
حَسْبُهُمْ ۚ وَلَعَنَهُمُ اللَّهُ ۖ وَهُمْ عَذَابٌ مُّقِيمٌ ۝

Artinya: "Allah mengancam orang-orang munafik laki-laki dan perempuan dan orang-orang kafir dengan neraka Jahannam, mereka kekal di dalamnya. Cukuplah neraka itu bagi mereka, dan Allah melaknati mereka, dan bagi mereka azab yang kekal,"

Penjelasan mengenai larangan memberikan berita hoaks dapat dipahami melalui dua surah yang disebutkan, yaitu Surah al-Jumu'ah dan al-Munafiqun. Tafsir dari Sayid Ali Khamene'i, Syekh Nashir Makarim Syirazi, dan Syekh Ja'far

Subhani memberikan pandangan mendalam terkait konsep munafik yang berkaitan dengan kejujuran dan integritas.

Surah al-Jumu'ah dan al-Munafiqun mencerminkan larangan terhadap sikap munafik, yang merupakan perilaku menampakkan sesuatu yang berbeda dengan yang sebenarnya, atau dalam konteks modern, menyebarkan berita hoaks yang bertentangan dengan kebenaran. Munafik digambarkan sebagai serigala berbulu domba, menunjukkan bahwa penampilan luar bisa menipu orang lain, tetapi hakikatnya penuh dengan kebohongan.

Ayat yang menyebutkan, "Dan di antara manusia ada yang berkata, 'Kami beriman kepada Allah dan hari akhir,' padahal sesungguhnya mereka itu bukanlah orang-orang yang beriman," menggambarkan perilaku munafik yang menipu Allah, orang-orang yang beriman, dan bahkan diri mereka sendiri. Hal ini menunjukkan bahwa menyembunyikan kebenaran dan menampilkan sesuatu yang palsu adalah perilaku munafik. Larangan menyebarkan berita hoaks dapat dihubungkan dengan konsep munafik ini. Berita hoaks atau informasi palsu seringkali digunakan untuk menipu dan memanipulasi opini publik, menunjukkan sikap munafik yang bertentangan dengan prinsip kejujuran dan kebenaran. Oleh karena itu, umat Muslim diajarkan untuk menghindari perilaku munafik, termasuk menyebarkan berita yang tidak benar.

Ayat yang menyebutkan ancaman Allah terhadap orang-orang munafik dengan neraka Jahannam yang kekal menegaskan seriusnya pelanggaran terhadap kejujuran dan integritas. Allah tidak hanya melaknati mereka, tetapi juga menjanjikan azab yang kekal. Ini memberikan penegasan bahwa kejujuran adalah nilai fundamental dalam ajaran Islam, dan pelanggaran terhadap nilai ini akan

berhadapan dengan konsekuensi yang serius di akhirat. Dalam konteks modern, larangan menyebarkan berita hoaks dalam Islam dapat dipahami sebagai panggilan untuk menjaga kebenaran, integritas, dan keadilan dalam berkomunikasi. Menyebarkan informasi yang akurat dan memastikan kejujuran dalam setiap ungkapan adalah bagian dari upaya untuk menjauhkan diri dari perilaku munafik dan mendekatkan diri kepada nilai-nilai yang ditekankan dalam Al-Qur'an (Kemenag RI, 2022).

2.5 Kajian Topik Dengan Teori Pendukung

Pada penelitian berjudul “*Market Basket Analysis for a Supermarket based on Frequent Itemset Mining*”, ditemukan bahwa *market basket analysis* merupakan hal penting dari sistem analisis untuk sebuah perusahaan yang berfungsi memperbaiki cara penempatan suatu produk atau keputusan dengan efektif, agar strategi *marketing* yang diberikan dapat sesuai dengan apa yang disukai oleh konsumen sehingga dapat menghasilkan kepuasan pada konsumen (Annie dkk, 2012).

Association rule adalah salah satu implementasi dari *Market Basket Analysis*. *Association rule* memiliki beberapa metode, diantaranya Apriori dan *FP-growth*. *FP-growth* dan LCM merupakan metode yang paling sering digunakan terkait penelitian dengan *data mining*, sebab algoritma ini dinilai sebagai algoritma yang paling seusai. Namun, algoritma apriori memiliki kekurangan, pada iterasi algoritma dalam mengamati *database* wajib dilakukan tiap iterasi, artinya butuh waktu yang lebih lama apabila lebih banyak iterasi yang dilakukan. Akan tetapi, iterasi-iterasi yang telah dilakukan tersebut, dapat mereduksi jumlah kandidat yang

harus dihitung nilai *support*-nya dengan pemangkasan. Reduksi jumlah kandidat ini yang menjadi sebab mengapa iterasi semakin banyak dilakukan maka waktu yang dibutuhkan lebih lama (Annie dkk, 2012).

BAB III

METODE PENELITIAN

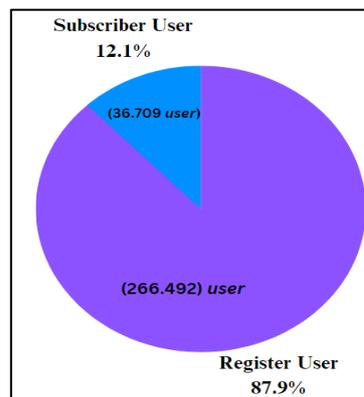
3.1 Jenis Penelitian

Penelitian kuantitatif adalah penelitian yang berfokus kepada fenomena yang objektif dan dianalisis dengan menggunakan metode kuantitatif. Untuk mencapai tingkat objektivitas yang maksimal, penelitian kuantitatif menggunakan angka-angka, pengolahan statistik, struktur, dan percobaan yang terkontrol (Sukmadinata, 2008). Jenis penelitian ini menggunakan pendekatan studi literatur dan metode kuantitatif. Studi literatur yang digunakan yaitu dengan memberikan referensi berdasarkan jurnal, artikel, dan buku yang memiliki relevansi untuk menyelesaikan penelitian ini. Penelitian kuantitatif menggunakan data berupa angka (Siyoto & Sodik, 2015).

3.2 Data dan Sumber Data

Variabel *dependent* (Y) pada penelitian ini merupakan *top category 1, top category 2, top category 3* pada 303201 *database* periode bulan Januari sampai September 2022. *Top category* adalah artikel dan atau berita yang dibaca oleh *user*, tiap *user* memiliki minimal satu artikel dan maksimal 3 top kategori artikel yang dibaca. Jenis berita dan atau artikel yang terdapat pada Kompas.id diantaranya di balik berita yang berisi tentang cerita di balik perjalanan pembuatan berita di Kompas.id, berita ekonomi yang membahas perkembangan aspek ekonomi di suatu negara/wilayah, artikel/berita *English* yang merupakan terjemahan seluruh artikel/berita Kompas.id dalam bahasa inggris, artikel gaya hidup yang membahas

aspek kehidupan sehari-hari serta kebiasaan, artikel hiburan yang membahas film, media, serta musik, sedangkan artikel humaniora membahas tentang pemahaman tentang manusia, budaya dan masyarakat, artikel ilmiah populer membahas tentang pengetahuan ilmiah dengan bahasa yang mudah dipahami orang awam, artikel internasional membahas tentang topik yang melibatkan hubungan antarnegara, isu global, politik internasional, ekonomi global, artikel muda membahas tentang tren yang dialami anak muda, kehidupan sekolah dan perkuliahan, artikel nusantara membahas tentang aspek yang berkaitan dengan wilayah Indonesia, artikel opini membahas tentang opini penulis Kompas tentang isu yang sedang marak di Indonesia, berita politik berisi tentang kebijakan yang terkait pada ranah politik dan hukum di Indonesia dan negara lain, riset, dan lain-lain. Data dapat dilihat pada Lampiran 1. Pada data tersebut terdapat 2 jenis *user* diantaranya *register user* dan *subscriber user*. *Register user* adalah *user* yang telah berlangganan di Kompas.id sedangkan *subscriber user* adalah *user* yang belum berlangganan di Kompas.id. Pada penelitian ini akan difokuskan pada *register user* saja, berikut merupakan gambaran jumlah tipe *register user* di Kompas.id.



Gambar 3.1 Pie Chart Tipe User Kompas.id

Berdasarkan Gambar 3.1 ditunjukkan pada *pie chart* pada bulan Januari hingga September 2022 terdapat 303201 jumlah *user* di Kompas.id. Diantaranya terdapat 36709 *subscriber user* yang memiliki persentase sebesar 12.1% dan *register user* sebanyak 266492 *user* dengan persentase sebesar 87.9%. Untuk meningkatkan jumlah pelanggan yang berlangganan atau *register user* maka kita hanya akan berfokus kepada 266492 *user* saja. Variabel X pada penelitian ini adalah jenis berita lainnya diantaranya akhir pekan, arsip, buku, daerah, di balik berita, diaspora, ekonomi, *English*, foto, gaya hidup, hiburan, humaniora, ilmiah populer, internasional, investigasi, kampus, lain-lain, metro, muda, multimedia, musik, nusantara, olahraga, opini, paparan topik, perjalanan, polhuk, riset, sastra, sekilas, tokoh, video. Berikut merupakan jumlah masing-masing artikel/berita yang menjadi *top category user*.

Tabel 3.1 Jumlah Artikel dan Opini Yang Dibaca *User*

Opini	Humaniora	Polhuk	Nusantara	Ekonomi
37902	2048	1582	1244	1193
Internasional	Video	Riset	Metro	Foto
989	842	764	543	521

Berdasarkan data terdapat 37902 *user* yang membaca artikel opini, yang mana merupakan artikel yang paling banyak dibaca oleh *user*. Sedangkan foto menjadi artikel yang paling sedikit dibaca oleh *user*, yaitu sebesar 521 pembaca saja. Dalam konteks ini, akan dicari relasi antar jenis berita. Variabel dependen adalah jenis berita akan dicari relasinya, sedangkan *variabel* independen adalah jenis berita lainnya yang digunakan untuk mencari pola atau hubungan dengan jenis

berita yang menjadi *variabel* dependen. Dengan menggunakan *association rule mining*, dapat identifikasi aturan asosiasi antara jenis tersebut. Misalnya dapat ditemukan *rule* seperti, “Jika seseorang membaca artikel jenis opini, maka akan kecenderungan bahwa mereka juga akan tertarik dengan berita jenis polhuk.” atau “Jika seseorang membaca berita jenis polhuk, maka ada kecenderungan bahwa mereka juga akan tertarik dengan berita jenis akhir pekan. Dengan menganalisis aturan-aturan ini, dapat diidentifikasi dua jenis berita yang paling disukai atau memiliki hubungan yang kuat dalam dataset tersebut.

3.3 Teknik Pengumpulan Data

Pada penelitian ini dilakukan teknik pengumpulan data sekunder. Data sekunder merupakan data yang didapatkan berdasarkan dokumen grafis (catatan, rapat, notulen, tabel, dan lain-lain), foto-foto, rekaman video, film, benda-benda, dan hal lain yang dapat memperbanyak data primer (Siyoto & Sodik, 2015). Dalam teknik ini, data yang diambil tidak dikumpulkan khusus untuk penelitian, melainkan sudah ada dan tersimpan dalam suatu *database* Kompas.id. Data didapatkan saat penulis magang di Harian Kompas.

3.4 Instrumen Penelitian

Instrumen Penelitian yang digunakan dalam penelitian ini berupa *software google colaboratory* dengan bahasa pemrograman python serta bahasa pemrograman R dengan menggunakan *tools R studio*. *Google colaboratory* dapat diakses pada laman colab.research.google.com. Proses pembersihan data dilakukan

menggunakan *tools google colab*, sedangkan analisis data menggunakan *tools R studio*.

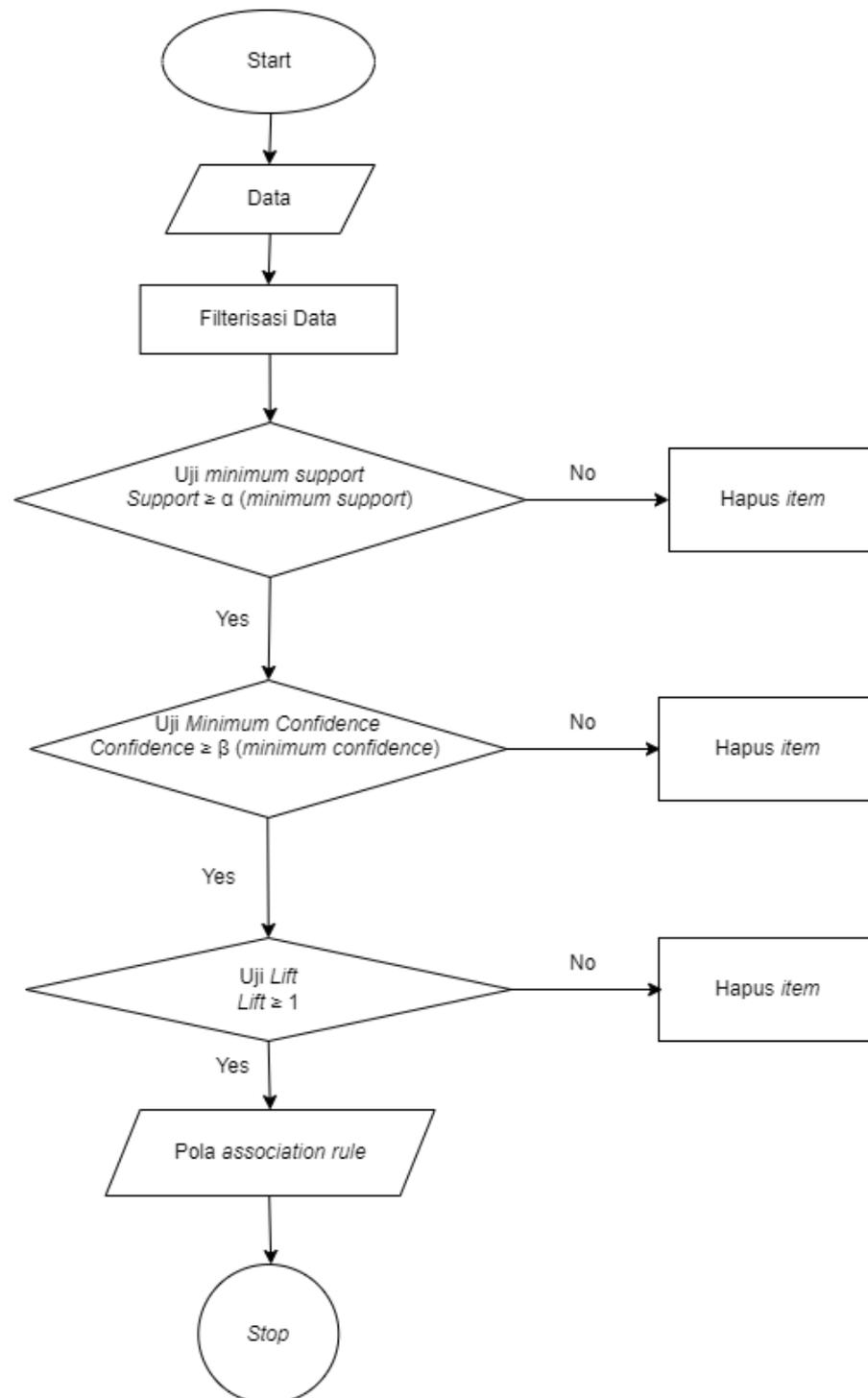
3.5 Teknik Analisis Data

Langkah yang dilakukan untuk melakukan penelitian ini adalah

1. Filterisasi data. Pada tahap ini dilakukan penghapusan data yang duplikat, menghapus data selain *register user*, dan juga menghapus data selain Kompas.id pada kolom '*source*'.
2. Pengujian *minimum support*. Tentukan nilai *minimum support*, hitung nilai *support* untuk tiap item dalam dataset. Periksa apakah nilai *support* sesuai dengan *minimum support*.
3. Pengujian *minimum confidence*. Tentukan nilai *minimum confidence*. Periksa apakah nilai *confidence* item lebih besar atau sama dengan nilai *minimum confidence* yang ditentukan.
4. Pengujian nilai *lift*. Hitung nilai *lift* untuk setiap aturan asosiasi yang dihasilkan. Tentukan apakah nilai *lift* menunjukkan hubungan yang signifikan atau tidak.
5. Pembuatan pola *association rule*. Ekstraksi aturan asosiasi dari dataset. Kemudian analisis aturan berdasarkan parameter seperti *support*, *confidence* dan *lift*.

3.6 Flowchart

Adapun *flowchart* dari Implementasi *data mining* yang dibuat adalah sebagai berikut:



Gambar 3.2 Flowchart

BAB IV
HASIL DAN PEMBAHASAN

4.1 Filterisasi Data

Pembersihan data pada penelitian ini dilakukan untuk menganalisis *register user* di Kompas.id. Pada database terdapat 303201 *users* dan 17 informasi terkait *users*. Variabel yang akan digunakan hanya *top_category_1*, *top_category_2*, dan *top_category_3* yang merupakan top artikel/berita yang sering dibaca/dibuka oleh *users*.

Terdapat 266392 *register user* atau *user* yang telah berlangganan di Kompas.id dan terdapat 36709 *subscriber user*. Setelah dilakukan pengecekan apakah terdapat data yang *duplicate*, dari data tersebut ditunjukkan bahwa tidak ada data yang *duplicate*. Selanjutnya, akan dilakukan filterisasi agar hanya *user* yang membaca artikel atau berita secara organik saja yang difokuskan. *User organik* disini merujuk pada *user* yang membaca berita/artikel tanpa terpengaruh oleh iklan atau voucher. Pada kolom “*source*” yang mencatat sumber link yang digunakan oleh *user* akan dibersihkan dari entry yang mengandung kata kunci seperti “voucher”, “artikel bebas akses”, dan “checkout.kompas.id/”. Kemudian, filterisasi data mendapatkan *organic user* dilakukan sebagai berikut:

Tabel 4.1 Filterisasi Organik *User*

id	Source	Source_platform	Source_hostname	Source_article_slug
	https://www.kompas.id/baca/ekonomi/2022/09/05/...	web	www.kompas.id	pekerja-upah-minimum-ikut-dapat-subsidi-upah

Lanjutan Tabel 4.1 Filterisasi Organik *User*

2	https://www.kompas.id/baca/utama/2018/08/25/bi...	web	www.kompas.id	binatang-politik
---	---	-----	---------------	------------------

Pengambilan data pada pada kolom *user_type* untuk mengambil data yang mengandung ‘*register*’ *users* dan pada kolom ‘*source*’ hanya akan diambil *link* utm yang mengandung “kompas.id/baca” untuk menyaring *link* utm yang tidak diinginkan pada data. Hal ini dilakukan karena *entry* tersebut menunjukkan bahwa *user* tersebut tidak masuk dalam kategori pengguna organik. Dengan melakukan pembersihan data ini, akan tercipta *subset* data yang khusus terdiri dari *register user* yang membaca artikel atau berita secara organik. *Source_article_slug* adalah judul *headline* berita/artikel.

Pencarian *frequent itemset* dilakukan dengan membaca *dataset* yang telah dibersihkan dan mengidentifikasi item-item (misalnya, produk atau layanan) yang paling sering muncul dalam transaksi.

Tabel 4.2 Frequent itemset

Opini	Humaniora	Polhuk	Nusantara	Ekonomi
37902	2048	1582	1244	1193
Internasional	Video	Riset	Metro	Foto
989	842	764	543	521

Pada Tabel 4.2 ditunjukkan jumlah *frequent itemset*, jumlah tertinggi ada pada artikel opini yaitu 37902 dan humaniora yaitu 2048 artikel. *Frequent itemset* ini akan membantu dalam penemuan item-item yang mungkin memiliki hubungan

kuat. Tahapan ini digunakan untuk menyortir itemset yang sering dan jarang muncul.

4.2 Pengujian Minimum Support (α)

Support adalah parameter yang mengukur sejauh mana item muncul bersamaan dengan item lain dalam transaksi. Dalam tahap ini dilakukan pengecekan nilai *support* untuk menentukan berapa nilai minimum *support*.

Tabel 4.3 Pemfilteran Minimum *Support*

No	LHS	RHS	<i>Support</i>
1	{foto}	{opini}	0.01374598
2	{opini}	{foto}	0.01374598
3	{metro}	{opini}	0.01432642
4	{opini}	{metro}	0.01432642
5	{riset}	{opini}	0.02015725
6	{opini}	{riset}	0.02015725

Dapat dilihat pada Tabel 4.3 bahwa nilai *support* ada di rentang 0.013 – 0.021, minimum *support* yang ditentukan adalah 0.021.

Tabel 4.4 Pembuatan *Support*

No	LHS	RHS	<i>Support</i>	Count
1	{humaniora}	{opini}	0.05403409	2048
2	{opini}	{humaniora}	0.05403409	2048
3	{polhuk}	{opini}	0.04173922	1582
4	{nusantara}	{opini}	0.03282149	1582

Pada Tabel 4.4 ditunjukkan bahwa minimum *support* sebesar 0.021. Kemudian akan diperiksa apakah nilai *support* item lebih besar atau sama dengan nilai minimum *support* yang telah ditentukan. *Itemset* yang tidak memenuhi nilai α akan dihapus dari dataset. Hal ini membantu mengurangi jumlah item yang kurang relevan dalam analisis. *Item* yang dihapus terlampir pada Tabel 4.3. Kemudian, *item* yang memenuhi minimum *support* akan dimasukkan ke *frequent* item.

Item yang memenuhi α dimasukkan ke frequent itemset, hal ini menandakan bahwa item tersebut memenuhi syarat untuk analisis lebih lanjut. Uji minimum *support* dilakukan dalam analisis market basket analisis untuk mengevaluasi sebesarnya besar atau kecil nilai *support* yang dapat digunakan agar aturan asosiasi yang dihasilkan masih dianggap relevan/bermakna.

Dalam pembentukan *association rules* berdasarkan nilai *support* ini, nilai *support* menjadi 49 faktor utama dalam menentukan keberadaan suatu *association rule*. Nilai *support* menggambarkan seberapa sering itemset di sebelah kiri (left hand side) dan itemset di sebelah kanan (right hand side) muncul bersamaan dalam dataset. Semakin tinggi nilai *support*, semakin sering pola asosiasi ditemukan dalam dataset. Nilai *support* yang tinggi menunjukkan adanya hubungan yang kuat antara itemset-itemset tersebut.

Dengan mengacu pada teori pembentukan *association rules*, analisis nilai *support* dapat memberikan pandangan tentang tingkat kekuatan hubungan antara itemset-itemset dalam dataset. Nilai *support* yang tinggi menunjukkan kemungkinan terjadinya pola asosiasi yang lebih signifikan, yang dapat digunakan sebagai acuan dalam pengambilan keputusan atau rekomendasi.

Berdasarkan Tabel 4.4, nilai *support* tertinggi yaitu {humaniora} => {opini} dengan nilai *support* sebesar 0.054. Hal ini menunjukkan bahwa sebanyak 5.4% dari total bacaan dalam dataset memiliki kedua item “humaniora” dan “opini” secara bersamaan. Sehingga dapat disimpulkan bahwa terdapat hubungan yang kuat antara topik humaniora dan opini dalam konteks data yang sedang dianalisis. Kemudian, nilai *support* tertinggi kedua yaitu {opini} dan {humaniora} dengan nilai *support* sebesar 0.054 dengan jumlah 2048 *rule*. Pada *association rule* dengan *support* terendah dilihat dari *data head* adalah {Nusantara} => {Opini} dengan nilai *support* sebesar 0.032. Hal ini menunjukkan bahwa sebanyak 3.2% dari total bacaan dalam dataset memiliki kedua item “nusantara” dan “opini” secara bersamaan. Berdasarkan dari data dengan *support* tertinggi dan terendah tidak terlalu memiliki selisih yang banyak karena kombinasi yang didapatkan cukup terbatas. Dalam contoh ini, ditemukan bahwa topik-topik seperti humaniora, polhuk dan nusantara memiliki keterkaitan yang signifikan dengan opini dalam data yang sedang dianalisis. Informasi ini dapat digunakan untuk pemahaman lebih lanjut tentang tren atau pola-pola dalam dataset dan dapat membantu dalam pengambilan keputusan atau perencanaan strategi.

4.3 Pengujian Nilai Minimum *Confidence*

Menentukan nilai β dapat disesuaikan dengan kebutuhan. Nilai *confidence* ditentukan sebesar 0.0541 sebab akan diambil nilai *confidence* tertinggi yaitu 1 dan nilai *confidence* terendah yaitu 0.05403409 akan dihapus dari *rule*.

Tabel 4.5 Pengujian Nilai Minimum *Confidence*

No.	LHS	RHS	<i>Confidence</i>	<i>Count</i>
1	{Humaniora}	{Opini}	1	2048
2	{Opini}	{Humaniora}	0.05403409	2048

Pengujian akan dilakukan untuk memeriksa apakah nilai *confidence* item lebih besar atau sama dengan nilai β yang telah ditentukan. Kemudian, dilakukan proses memasukkan item yang memenuhi minimum *confidence*. Setelah dilakukan penghapusan item yang tidak memenuhi β , item yang memenuhi uji β akan dihapus. Item yang memiliki nilai *confidence* yang memenuhi atau melebihi β akan dimasukkan ke dalam hasil analisis asosiasi. Ditunjukkan tabel pembentukan *association rule* berdasarkan salah satu parameter, yaitu nilai *confidence*:

Tabel 4.6 Pembentukan *Association Rule* Berdasarkan Nilai *Confidence*

No.	LHS	RHS	<i>Confidence</i>	<i>Count</i>
1	{Ekonomi} →	{Opini}	1	1193
2	{Nusantara} →	{Opini}	1	1244
3	{Polhuk} →	{Opini}	1	1582
4	{Humaniora} →	{Opini}	1	2048
5	{Opini} →	{Humaniora}	1	2048

Pada pembentukan *association rule* seluruh rule memiliki nilai tertinggi pada *confidence*-nya, yaitu sebesar 100% atau 1, ini menunjukkan bahwa ketika LHS muncul maka nilai RHS juga akan muncul dan sebaliknya. Dapat disimpulkan pada Tabel 4.6 memiliki rule dengan korelasi yang kuat.

4.4 Pengujian Nilai *Lift*

Cara mendapatkan nilai *lift* dijelaskan pada Persamaan (2.4) *Lift* yang lebih besar dari 1 menunjukkan adanya hubungan yang positif antara LHS dan RHS. *Lift* sebesar 1 menunjukkan hubungan yang netral, sedangkan *lift* kurang dari 1 menunjukkan hubungan yang negatif atau saling mengecualikan (Rerung, 2018). Dalam tahap ini dilakukan pemeriksaan apakah nilai *lift* untuk subset yang dihasilkan dari itemset tersebut lebih besar atau sama dengan 1. Jika *lift* kurang dari 1, itu menunjukkan bahwa asosiasi tersebut lebih lemah dari yang diharapkan secara acak dan subset tersebut dapat dihapus. Ditunjukkan pada Tabel 4.7 bahwa nilai *lift* seluruhnya memiliki nilai sebesar 1.

Tabel 4.7 Penentuan nilai *lift*

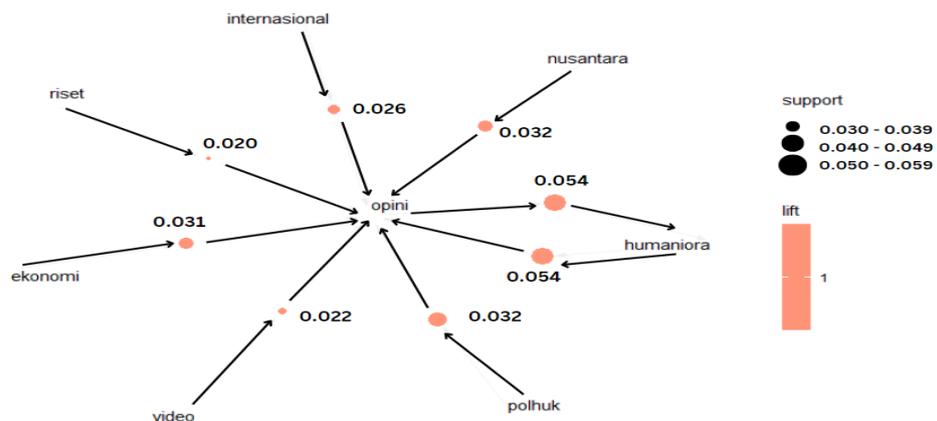
No.	LHS	RHS	<i>Lift</i>
1	{Ekonomi} →	{Opini}	1
2	{Nusantara} →	{Opini}	1
3	{Polhuk} →	{Opini}	1
4	{Humaniora} →	{Opini}	1

Lanjutan Tabel 4.7 Penentuan nilai *lift*

5	{Opini} →	{Humaniora}	1
---	-----------	-------------	---

4.5 Pembuatan Pola *Association Rule*

Pada langkah ini dapat dihasilkan pola aturan asosiasi berdasarkan item-item yang telah melewati semua uji dan kriteria sebelumnya. Pola aturan ini digunakan untuk mendapatkan wawasan tentang hubungan antara item-item dalam data transaksi. Pola *association rule* adalah salah satu teknik analisis data yang digunakan untuk mengungkapkan hubungan dan keterkaitan antara item-item dalam sebuah dataset. Dalam pola *association rule*, item-item dari dataset diwakili oleh simpul-simpul yang saling terhubung oleh garis-garis. Garis-garis ini mencerminkan hubungan antara item-item tersebut, yang ditentukan oleh tingkat kekuatan atau kepercayaan (*support* dan *confidence*) dari aturan-aturan yang ditemukan. Berikut akan ditunjukkan hasil diagram *association rule* hasil dari penelitian *register user Kompas.id*.



Gambar 4.1 Diagram *Association Rule Register User Kompas.id*

Berdasarkan Gambar 4.1 menunjukkan bahwa artikel/berita polhuk, internasional, riset, video, nusantara, ekonomi, humaniora memiliki keterkaitan

yang kuat dengan artikel opini. Ditunjukkan dengan semakin besar ukuran lingkaran pada diagram maka semakin besar pula nilai *support* kedua itemset tersebut, dapat dilihat bahwa nilai *lift* pada itemset seluruhnya memiliki nilai 1. Dapat disimpulkan dari Gambar 4.1 bahwa nilai *support* antara {opini} => {humaniora} merupakan yang tertinggi, urutan selanjutnya adalah antara artikel {polhuk} dan {opini}. Terdapat aturan asosiasi berdasarkan *Support* Pada beberapa aturan asosiasi terdapat tingkat *support* yang tinggi. Sebagai contoh, aturan {Humaniora} => {Opini} memiliki tingkat *support* sebesar 0.05403409, yang berarti bahwa kombinasi antara artikel humaniora dan opini muncul dalam sekitar 5.4% dari total transaksi. Ini adalah aturan dengan *support* tertinggi yang menunjukkan adanya hubungan kuat antara kedua jenis artikel/berita tersebut.

Sedangkan aturan asosiasi berdasarkan *confidence*, *confidence* mengukur seberapa sering konsekuensi (rhs) terjadi ketika antecedent (lhs) terjadi. Aturan dengan *confidence* tinggi menunjukkan hubungan yang kuat antara antecedent dan konsekuensi. Sebagai contoh, aturan {Internasional} => {Opini} memiliki *confidence* sebesar 1.0, yang berarti bahwa setiap kali artikel “Internasional” muncul, artikel “Opini” selalu ikut muncul. Ini menunjukkan hubungan yang kuat antara kedua jenis artikel tersebut.

Sedangkan aturan asosiasi berdasarkan *lift*, *lift* mengukur sebesar besar hubungan antara antecedent (LHS) dan RHS dibandingkan jika dengan jika keduanya terjadi secara independent. Aturan dengan *lift* tertinggi menunjukkan hubungan yang paling kuat. Sebagai contoh, aturan {Foto} => {Opini} memiliki *lift* sebesar 1, yang menunjukkan bahwa hubungan antara Foto dan Opini adalah hubungan yang kuat.

Hasil analisis market basket analysis dengan algoritma Apriori menunjukkan adanya pola asosiasi kuat antara jenis artikel/berita tertentu. Aturan dengan *support*, *confidence*, dan *lift* tinggi dapat digunakan untuk mengidentifikasi hubungan yang signifikan antara artikel/berita, yang dapat berguna dalam merancang strategi pemasaran atau merekomendasikan artikel kepada pembaca.

4.6 Kajian Integrasi Topik Dalam Islam

Allah SWT berfirman pada surat Al-Qamar ayat 52:

وَكُلُّ شَيْءٍ فَعَلُوهُ فِي الزُّبُرِ ﴿٥٢﴾

Artinya: “Dan segala sesuatu yang telah mereka perbuat tercatat dalam buku-buku catatan.”

Menurut tafsir Tahlili, semua perbuatan manusia terhimpun dalam buku catatan masing-masing. Hal itu karena setiap perbuatan manusia kecil ataupun besar, baik atau buruk dicatat oleh malaikat di dalam buku catatan itu, Malaikat Kiram atau sebut saja Raqib, mencatat perbuatan yang baik dan Malaikat Katibin atau 'Atid mencatat perbuatan yang tidak baik. Dalam ayat lain Allah berfirman: Tidak ada suatu kata yang diucapkannya melainkan ada di sisinya malaikat pengawas yang selalu siap (mencatat). Oleh karena semua aktivitas manusia, baik perbuatan maupun ucapan, baik yang baik maupun yang buruk, baik besar maupun kecil tercatat di dalam buku catatan masing-masing, maka sangat mudah bagi Allah menjatuhkan hukuman kepada yang berdosa dan memberikan pahala kepada yang berbuat baik.

Penerapan data mining dengan teknik *association rule* pada konsep pencatatan perbuatan manusia oleh malaikat dalam buku catatan masing-masing dapat memberikan wawasan yang lebih mendalam tentang hubungan antara

perbuatan baik atau buruk dengan kecenderungan perbuatan atau ucapan lainnya. Dalam konteks ini, setiap perbuatan manusia dianggap sebagai suatu entitas atau item yang dicatat dalam transaksi, dan buku catatan masing-masing dianggap sebagai kumpulan transaksi. Teknik *association rule* membantu mengidentifikasi pola atau asosiasi antaritem yang mungkin sulit dilihat secara langsung.

Dengan menerapkan data mining, kita dapat menemukan aturan asosiasi yang menggambarkan hubungan antara perbuatan manusia. Misalnya, dapat ditemukan aturan seperti: "Jika seseorang melakukan perbuatan baik A, maka kemungkinan besar dia juga akan melakukan perbuatan baik B." Atau sebaliknya, "Jika seseorang melakukan perbuatan buruk X, maka kemungkinan besar dia juga akan melakukan perbuatan buruk Y." Aturan-aturan ini mencerminkan pola atau hubungan yang ada dalam data pencatatan malaikat.

Penerapan data mining pada konsep ini dapat membantu memahami lebih baik dinamika perbuatan manusia dan bagaimana perbuatan-perbuatan tersebut saling terkait. Analisis ini dapat memberikan pandangan yang lebih rinci tentang faktor-faktor yang mempengaruhi perbuatan baik atau buruk seseorang. Selain itu, dengan pemahaman yang lebih mendalam tentang asosiasi antarperbuatan, kita dapat menggali pengetahuan yang mungkin tidak terlihat secara kasat mata.

Dalam konteks agama dan ayat yang disebutkan, data mining dapat memberikan ilustrasi tentang bagaimana Allah dengan mudah dapat mengetahui perbuatan manusia dan memberikan konsekuensi yang sesuai. Dengan memahami pola-pola ini, kita dapat melihat bagaimana sistem pencatatan oleh malaikat memberikan dasar bagi konsep pahala dan hukuman. Penerapan data mining pada

konsep agama ini juga dapat diartikan sebagai upaya untuk mendekati pemahaman ilmiah terhadap konsep-konsep keagamaan yang kompleks.

Selain itu, penerapan data mining dalam konteks pencatatan oleh malaikat juga dapat memiliki implikasi praktis. Misalnya, dengan memahami pola asosiasi antarperbuatan, kita dapat merancang strategi atau sistem yang lebih efektif untuk membimbing manusia dalam mencapai perbuatan baik dan menghindari perbuatan buruk. Dengan analisis yang cermat, mungkin kita dapat mengidentifikasi faktor-faktor pemicu yang mempengaruhi seseorang untuk melakukan perbuatan tertentu. Sebagai contoh, apakah terdapat pola khusus yang memicu perbuatan baik pada individu tertentu atau sebaliknya.

Selain itu, penerapan data mining pada konsep pencatatan malaikat juga memunculkan pertanyaan etis terkait privasi dan kebebasan individu. Pemahaman mendalam tentang perilaku manusia melalui analisis data mining dapat membuka pintu untuk pertimbangan etis terkait hak privasi dan otonomi individu. Bagaimana data ini digunakan dan apakah pengungkapan informasi pribadi dapat diterima atau tidak, menjadi pertimbangan penting dalam mengembangkan pendekatan yang sejalan dengan nilai-nilai kemanusiaan.

Dengan demikian, penerapan data mining pada konsep agama membuka peluang untuk mengintegrasikan aspek keilmuan dan etika dalam membahas prinsip-prinsip agama. Sementara konsep agama sendiri memberikan kerangka kerja moral, data mining memberikan alat untuk menganalisis, memahami, dan menerapkan konsep tersebut secara lebih kontekstual dalam kehidupan sehari-hari.

BAB V

KESIMPULAN

5.1 Kesimpulan

Pola *association rule* artikel/berita yang dihasilkan dari penerapan analisis market basket analisis dengan menggunakan algoritma apriori pada *register user* Kompas.id yaitu terdapat beberapa pola *association rule* yang memiliki nilai *support* dan *confidence* yang cukup tinggi. Sebagai contoh, aturan yang menyatakan bahwa artikel {Humaniora} memiliki kecenderungan untuk memiliki artikel {Opini} dan sebaliknya. Kemudian, topik opini menjadi topik yang terkait dengan berbagai topik berita lainnya. Meskipun nilai *support* dan *confidence* berbeda-beda, namun semua aturan menunjukkan kecenderungan artikel dengan topik tertentu untuk juga memiliki topik {Opini}.

5.2 Saran

Berdasarkan hasil pola *association rule* maka diperoleh saran sebagai berikut:

1. Kepada Perusahaan Kompas: Hasil analisis ini dapat digunakan untuk merancang strategi pemasaran yang lebih efektif. Misalnya, jika kita tahu bahwa artikel Humaniora dan Opini memiliki hubungan yang kuat, kita dapat mempertimbangkan untuk menggabungkan atau mengkategorikan artikel tersebut bersama dalam promosi atau rekomendasi kepada pembaca.
2. Rekomendasi Artikel: Analisis ini juga dapat digunakan untuk membuat rekomendasi artikel kepada pembaca. Jika seorang pembaca membaca

artikel Foto, berdasarkan aturan dengan *confidence* tinggi, kita dapat merekomendasikan artikel Opini kepada mereka, karena kemungkinan besar mereka juga akan tertarik pada jenis artikel tersebut. Dapat disimpulkan bahwa analisis market basket analysis dengan algoritma Apriori telah memberikan wawasan yang berharga tentang pola asosiasi antara jenis artikel/berita, yang dapat digunakan untuk mendukung keputusan pemasaran dan rekomendasi kepada pembaca. Analisis semacam ini dapat membantu meningkatkan efektivitas strategi pemasaran dan meningkatkan pengalaman pembaca.

DAFTAR PUSTAKA

- Agrawal, R. and Srikant, R. (1994) Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the 20th International Conference on Very Large Data Bases, Santiago de Chile, 12-15 September 1994, 487-499.
- Annie, L.C., & Ashok Kumar, D. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining. IJCSI International Journal of Computer Science Issues, ISSN (Online): 1694-0814.
- Ahmad, Abu Hania. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. Yayasan Cahaya Islam, Jurnal Teknologi Indonesia.
- Bickel, W. K., dkk. (2023). Predictors of smoking cessation outcomes identified by machine learning: A systematic review. *Addiction neuroscience*, 6, 100068. <https://doi.org/10.1016/j.addicn.2023.100068>
- Bramer, M. (2007). Principles of Data Mining. London: Springer
- Dewi, H. F., Handayani, H. H., & Indra, J. (2022). Implementasi Algoritma Apriori Terhadap Market Basket Analysis Pada Data Penjualan Retail. *JINTEKS*, vol. 4, no. 4, hal. 432-436
- Fauzy, Mohamad & Kemas, Rahmat & Asror, Ibnu. (2016). Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung. *Jitter*. II. 221. 10.33197/jitter.vol2.iss3.2016.111.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230>
- Gillis, Alexander S. (2022). What is an algorithm? | TechTarget. Diambil dari <https://www.techtarget.com/whatis/definition/algorithm>
- Gonrunescu, F. (2011). *Data Mining* (1st ed.). London: Springer
- Hirji, K. K. (2001). Exploring Data Mining Implementation. *Communications of the ACM*. 44(7). 10.1145/379300.379323
- Husna, Radhiatul & Lestari, Riri & Hendra, Yomei. (2019). Inventory model of goods availability with apriori algorithm. *Journal of Physics: Conference Series*. 1317. 012019. 10.1088/1742-6596/1317/1/012019.
- Kemenag RI. (2023). *Qur'an Kemenag*. Lajnah Pentashihan Mushaf Al-Qur'an.
- Kementerian Agama RI, (2012). *Al-Qur'an dan Terjemahnya*. Bandung: Syamil Qur'an
- Kesavaraj, G., & Sukumaran, S. (2013). A study on classification techniques in data mining. In 2013 Fourth International Conference on Computing,

- Communications and Networking Technologies (ICCCNT). 1-7. 10.1109/ICCCNT.2013.6726842.
- Kim, Hyea & Kim, Jae Kyeong & Chen, Qiu. (2012). A product network analysis for extending the market basket analysis. *Expert Systems with Applications*. 39. 7403–7410. 10.1016/j.eswa.2012.01.066.
- Kompas. (2018). Kompas.
- Lin, Hsiao-Kang & Hsieh, Cheng-Huan & Wei, Nai-Chieh & Peng, Yi-Chun. (2019). Association rules mining in R for product performance management in industry 4.0. *Procedia CIRP*. 83. 699-704. 10.1016/j.procir.2019.04.099.
- Majjah, Ibnu. (2016). *Jual Beli Berdasarkan Suka Sama Suka*.
- Poore, T. & Taylor-Cousar, Jennifer & Zemanick, Edith. (2021). Cardiovascular complications in cystic fibrosis: A review of the literature. *Journal of Cystic Fibrosis*. 21. 10.1016/j.jcf.2021.04.016.
- Rerung, Rintho. (2018). Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk. *Jurnal Teknologi Rekayasa*. 3. 89. 10.31544/jtera.v3.i1.2018.89-98.
- Santoso, M.H. (2021). Application of *Association Rule* Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom. *Brilliance: Research of Artificial Intelligence*. 1(2), 54-66. <https://doi.org/10.47709/brilliance.v1i2.1228>
- Shevira, S., Suarjaya, I., & Buana, P. (2022). Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia. *JITTER : Jurnal Ilmiah Teknologi Dan Komputer*, 3(2), 1074-1081. doi:10.24843/JTRTI.2022.v03.i02.p06
- Silva, Jesus & Varela, Noel & López, Luz & Rojas Millan, Rafael. (2019). Association Rules Extraction for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm. *Procedia Computer Science*. 151. 1207-1212. 10.1016/j.procs.2019.04.173.
- Siyoto, S., & Sodik, M. A. (2015). *Dasar Metodologi Penelitian*. Sleman: Literasi Media
- Son, L.H., Chiclana, F., Kumar, R., Mittal, M., Khari, M., Chatterjee, J.M., & Baik, S.W. (2018). ARM-AMO: An efficient association rule mining algorithm based on animal migration optimization. *Knowl. Based Syst.*, 154, 68-80.
- Sukmadinata, N. S. (2008). *Metode Penelitian Pendidikan*. Bandung: Remaja Rosdakarya
- Suyanto. (2019). *Data Mining: Untuk Klasifikasi Dan Klasterisasi Data*. Bandung: Penerbit Informatika
- Syamsudin, M. (2021). *Surat Al-Baqarah Ayat 282 dan Sifat Pasar Berjangka*.

- Zanuardi, Arvian & Suprayitno, Hitapriya. (2018). Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan Knowledge Discovery in Database. *Jurnal Manajemen Aset Infrastruktur & Fasilitas*. 2. 10.12962/j26151847.v2i1.3767.
- Zhang, Chaobo & Zhao, Yang & Zhang, Xuejun. (2019). An improved association rule mining-based method for discovering abnormal operation patterns of HVAC systems. *Energy Procedia*. 158. 2701-2706. 10.1016/j.egypro.2019.02.025.
- Zhang, Chengqi & Zhang, Shichao. (2002). Association Rule Mining, Models and Algorithms. 10.1007/3-540-46027-6.

LAMPIRAN

Transaksi	top_category_1	top_category_2	top_category_3
1	ekonomi	english	
2	polhuk		
3	polhuk	video	investigasi
4	english		
5	opini	gaya-hidup	
6	opini		
7	opini		
8	foto		
9	video	opini	humaniora
10	buku	opini	
11	opini	nusantara	video
12	sastra		
13	opini		
14	humaniora	opini	
15	internasional		
16	nusantara	ekonomi	
17	opini	ekonomi	foto
18	opini		
19	nusantara		
20	opini	ekonomi	humaniora
21	polhuk	humaniora	internasional
22	olahraga		
23	internasional	polhuk	
24	polhuk	humaniora	
25	opini	english	polhuk
26	sastra		
27	polhuk		
28	internasional	riset	
29	polhuk	opini	nusantara

RIWAYAT HIDUP



Aulia Gita Pratiwi, lahir di Madiun pada 19 Januari 2001. Penulis merupakan anak kedua dari 2 bersaudara. Penulis telah menempuh pendidikan mulai dari TK Masyithoh Madiun dan lulus pada tahun 2007, dilanjutkan menempuh pendidikan sekolah dasar di SDN 03 Madiun Lor dan lulus pada tahun 2013. Kemudian penulis melanjutkan pendidikan sekolah menengah pertama di SMPN 3 Madiun dan lulus pada tahun 2016. Selanjutnya menempuh pendidikan sekolah menengah atas di SMAN 5 Madiun dan lulus pada tahun 2019. Pada tahun yang sama, penulis melanjutkan pendidikan di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada Program Studi Matematika, Fakultas Sains dan Teknologi.

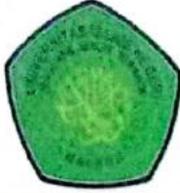
Selama menempuh pendidikan di perguruan tinggi, penulis aktif mengikuti magang di Harian Kompas, bekerja sebagai Strategic Planner, Content Creator, AI Engineer, Data Analyst dan mengikuti kegiatan relawan di luar kampus.



BUKTI KONSULTASI SKRIPSI

Nama : Aulia Gita Pratiwi
NIM : 19610057
Fakultas/Program Studi : Sains dan Teknologi/Matematika
Judul Skripsi : Implementasi Data Mining Menggunakan Metode Association Rule Pada Register User Kompas.id
Pembimbing I : Angga Dwi Mulyanto, M.Si.
Pembimbing II : Erna Herawati, M.Pd.

No	Tanggal	Hal	Tanda Tangan
1.	7 Oktober 2022	Konsultasi Judul dan Topik	1.
2	23 Desember 2022	Konsultasi BAB I, BAB II, dan BAB III	2.
3	24 Desember 2022	Konsultasi Keagamaan BAB I, BAB II	3.
4	10 Januari 2023	Konsultasi Revisi BAB I, BAB II, dan BAB III	4.
5	12 April 2023	ACC Proposal Skripsi	5.
6	17 April 2023	Konsultasi Revisi Keagamaan BAB I, BAB II	6.
7	31 Mei 2023	Konsultasi BAB IV dan BAB V	7.
8	5 Juni 2023	Konsultasi Revisi BAB IV dan BAB V	8.
9	8 Juli 2023	ACC Seminar Hasil	9.
10	13 September 2023	Konsultasi Keagamaan BAB IV	10.
11	21 September 2023	Konsultasi Revisi Pertama Keagamaan BAB IV	11.
12	3 November 2023	Konsultasi Revisi Kedua Keagamaan BAB IV	12.
13	25 November 2023	Konsultasi Revisi Seminar Hasil	13.
14	7 Desember 2023	ACC Sidang Skripsi	14.
15	18 Desember 2023	Revisi Abstrak Arab	15.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

16	19 Desember 2023	ACC Abstrak Arab	16.
17.	20 Desember 2023	ACC Keagamaan	17.
18	20 Februari 2024	ACC Keseluruhan	18.

Malang, 20 Februari 2024

Mengetahui,

Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.

NIP. 19741129 200012 2 005