

**KLASIFIKASI DATA *SCIMAGO JOURNAL* DAN *COUNTRY RANK*  
MENGUNAKAN ALGORITMA C4.5**

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Jurusan Statistika**



**Disusun Oleh:**

**Erdwika Putri**

**(14611219)**

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2018**

**KLASIFIKASI DATA *SCIMAGO JOURNAL* DAN *COUNTRY RANK*  
MENGUNAKAN ALGORITMA C4.5**

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Jurusan Statistika**



**Disusun Oleh:**

**Erdwika Putri**

**(14611219)**

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2018**

**HALAMAN PERSETUJUAN DOSEN PEMBIMBING**

**TUGAS AKHIR**

Judul : *Klasifikasi Data Scimago Journal Dan Country Rank Menggunakan Algoritma C4.5*  
Nama Mahasiswa : Erdwika Putri  
Nomor Mahasiswa : 14 611 219

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK  
DIUJIKAN**

**Yogyakarta, April 2018**

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ  
الْحَمْدُ لِلَّهِ الَّذِي هَدَانَا لِهَذَا وَمَا كُنَّا لِنَشْكُرَهُ لَوْلَا رَحْمَتُ اللَّهِ عَلَيْنَا لَكُنَّا مِنَ الْخَاسِرِينَ

**Pembimbing**

**(Ayundyah Kesumawati, S.Si., M.Si.)**

HALAMAN PENGESAHAN  
TUGAS AKHIR

KLASIFIKASI DATA *SCIMAGO JOURNAL* DAN *COUNTRY RANK*  
MENGUNAKAN ALGORITMA C4.5

Nama Mahasiswa : Erdwika Putri  
Nomor Mahasiswa : 14 611 2119

TUGAS AKHIR INI TELAH DIUJIKAN  
PADA TANGGAL, April 2018

Nama Penguji

Tanda Tangan

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

2. Muhammad Muhajir, S.Si., M.Sc

3. Ayundyah Kesumawati, S.Si., M.Si

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Drs. Alwar, M.Sc., Ph.D)

## KATA PENGANTAR



*Assalamualaikum Wr. Wb*

Alhamdulillahirobbil alamin, Puji dan syukur kehadiran Allah SWT atas berkat rahmat serta kasih-Nya yang selalu dilimpahkan kepada penulis, sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan judul “Klasifikasi Data *Scimago Journal* Dan *Country Rank* Menggunakan Algoritma C4.5”. Shalawat beserta salam semoga senantiasa terlimpah curahkan kepada Nabi Muhammad SAW, kepada keluarganya, para sahabatnya, hingga kepada umatnya hingga akhir zaman.

Penulisan Tugas Akhir ini diajukan untuk memenuhi salah satu syarat untuk memperoleh gelar Sarjana Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. terselesaikannya skripsi ini tidak terlepas dari bantuan banyak pihak, sehingga pada kesempatan ini dengan segala kerendahan hati dan penuh rasa hormat, penulis menghaturkan terima kasih yang sebesar-besarnya bagi semua pihak yang telah memberikan bantuan moril maupun materil, baik langsung maupun tidak langsung dalam penyusunan Tugas Akhir ini hingga selesai, terutama kepada :

1. Bapak Drs. Allwar, M.Sc., Ph.D selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta beserta seluruh jajarannya.
2. Bapak Dr. Fajriya Hakim, S.Si., M.Si selaku Ketua Program Studi Statistika beserta seluruh jajarannya.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si selaku Dosen Pembimbing penulis yang telah memberikan kritik dan saran bimbingan maupun arahan yang sangat berguna dalam penyusunan Tugas Akhir ini.
4. Dosen-dosen Statistika Universitas Islam Indonesia yang telah mendidik dan berbagi ilmunya kepada penulis.

5. Kedua orang tua tercinta, Bapak Lahidin dan Ibu Warti yang selalu memberikan semangat dan dukungan tiada henti, serta selalu mengiringi lewat untaian doa disetiap langkah penulis.
6. Kakak penulis, Sri Rahayu Putri yang selalu sabar dan mengerti setiap keluh kesah yang penulis rasakan, serta adik penulis Ghulam Akbar yang selalu memberikan semangat yang membara bagi penulis.
7. Sahabat seperjuangan Een Juliani yang selalu berbagi keluh kesah, curahan hati, sekaligus memberikan semangat, doa dan motivasi yang luar biasa dalam pengerjaan Tugas Akhir ini.
8. Sahabat penulis Anisa, Ayu, Dhea, Ditia, Khusnul, Maulida, Raby, Reny, Sari, Sita dan Zarina yang selama ini selalu belajar dan berjuang bersama dalam mengerjakan Tugas Akhir ini.
9. Manna Squad Wadel, Shinta, Rian, Oddy, dan Dimas yang selalu memberikan kebahagiaan yang luar biasa dikala rindu akan kampung halaman.
10. Teman- Teman satu bimbingan tugas akhir Sari, Maulida, Suci, Herlina, Riza, Afifah, Reny, Ayu, Molydah, Rara, Elsa, Dian, Purwina, Galih, Yayan, Achmad, dan Syauqi yang senantiasa selalu berbagi ilmu, cerita dan pengalaman.
11. Sahabat Statistika 2014 (XIX) yang sudah banyak memberikan semangat kepada penulis dalam penyusunan Tugas Akhir ini.
12. Teman-teman KKN unit 118 Desa Cacaban Kidul, Kecamatan Bener, Kabupaten Purworejo, Hanifah, Anggita, Ana, Syifa, Bagas, Latif, Viko, dan Adek.
13. Semua pihak yang telah membantu dalam penyelesaian Tugas Akhir ini, yang tidak bisa penulis sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, oleh karena itu penulis mengharapkan kritik dan saran yang bersifat membangun dari semua pihak demi keempurnaan Tugas Akhir ini. Penulis berharap semoga Tugas Akhir ini dapat memberikan manfaat bagi penulis dan bagi semua yang

membutuhkan. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua, Aamiin aamiin ya robbal'amin.

*Wassalamualaikum, Wr.Wb*

Yogyakarta, April 2018



Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN .....	<b>Error! Bookmark not defined.</b>
HALAMAN PENGESAHAN .....	<b>Error! Bookmark not defined.</b>
KATA PENGANTAR .....	iii
DAFTAR ISI .....	vii
DAFTAR TABEL .....	ix
DAFTAR GAMBAR .....	x
DAFTAR LAMPIRAN .....	xi
PERNYATAAN .....	<b>Error! Bookmark not defined.</b>
INTISARI .....	xiii
ABSTRACT .....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Penelitian .....	4
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	5
BAB II TINJAUAN PUSTAKA .....	6
BAB III LANDASAN TEORI .....	13
3.1 Jurnal Ilmiah .....	13
3.2 <i>Scimago Journal &amp; Country Rank</i> .....	15
3.3 <i>Data Mining</i> .....	16
3.4 Klasifikasi .....	17
3.5 <i>Data Training dan Data Testing</i> .....	19
3.6 <i>Pohon Keputusan (Decision Tree)</i> .....	19
3.7 Algoritma C4.5 .....	21
3.7.1 Pengertian Algoritma C4.5 .....	21
3.7.2 Langkah-langkah Algoritma C4.5 .....	22

3.7.3	Metode Pencarian Atribut Pada <i>Node</i> .....	23
3.7.4	Aturan Pemberhentian .....	24
3.7.5	<i>Confusion Matrix</i> .....	25
3.7.6	<i>K-fold Cross Validation</i> .....	26
3.8	Analisis Regresi .....	26
<b>BAB IV METODOLOGI PENELITIAN</b> .....		28
4.1	Populasi Penelitian .....	28
4.2	Jenis dan Sumber Data .....	28
4.3	Variabel Penelitian .....	28
4.4	Metode Pengumpulan Data .....	30
4.5	Metode Analisis Data .....	30
4.6	Tahapan Penelitian <i>Decision Tree</i> Metode Algoritma C4.5 .....	30
<b>BAB V HASIL DAN PEMBAHASAN</b> .....		33
5.1	Analisis Deskriptif .....	33
5.2	Analisis Regresi .....	39
5.3	Klasifikasi <i>Decision Tree</i> Dengan Algoritma C4.5 .....	42
5.3.1	Penentuan Atribut Pada <i>Node</i> .....	42
5.3.2	Pohon Keputusan Algoritma C4.5 .....	43
<b>BAB VI PENUTUP</b> .....		47
6.1	Kesimpulan .....	47
6.2	Saran .....	47
<b>DAFTAR PUSTAKA</b> .....		48
<b>LAMPIRAN</b> .....		53

## DAFTAR TABEL

<b>Tabel 2.1</b> Perbandingan Penelitian	8
<b>Tabel 3.1</b> Kategori Subjek Pada SJR	16
<b>Tabel 3.2</b> Frekuensi Penggunaan Algoritma Pohon Keputusan	20
<b>Tabel 3.3</b> Confusion Matrix 2 Kelas	25
<b>Tabel 4.1</b> Definisi Operasional Variabel	29
<b>Tabel 5.1</b> Uji Overall	39
<b>Tabel 5.2</b> Uji Parsial	41
<b>Tabel 5.3</b> Segmentasi Pohon Keputusan Algoritma C4.5	44
<b>Tabel 5.4</b> Perbandingan Nilai Akurasi Machine Learning	45
<b>Tabel 5.5</b> Confusion Matrix	46

## DAFTAR GAMBAR

<b>Gambar 3.1</b> Teknik Klasifikasi	18
<b>Gambar 3.2</b> Konsep Pohon Keputusan	19
<b>Gambar 3.3</b> Konsep Dasar Pohon Keputusan	20
<b>Gambar 3.4</b> Flowchart Algoritma C4.5	22
<b>Gambar 4.1</b> Flowchart Penelitian	31
<b>Gambar 5.1</b> Presentase Negara Pada Q1	33
<b>Gambar 5.2</b> Presentase Negara Pada Q2	34
<b>Gambar 5.3</b> Presentase Negara Pada Q3	35
<b>Gambar 5.4</b> Presentase Negara Pada Q4	36
<b>Gambar 5.5</b> Persebaran Negara Berdasarkan Nilai SJR	37
<b>Gambar 5.6</b> Presentase Kategori Jurnal	38
<b>Gambar 5.7</b> Perbandingan Kategori Berdasarkan Quartile	39
<b>Gambar 5.8</b> Prediksi Kelas	44

## DAFTAR LAMPIRAN

<b>Lampiran 1.</b> Dataset SJR	53
<b>Lampiran 2.</b> Syntax Decision Tree Algoritma C4.5	57
<b>Lampiran 3.</b> Hasil Klasifikasi	58

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan daftar pustaka.

Yogyakarta, April 2018



Penulis

# **KLASIFIKASI DATA *SCIMAGO JOURNAL* DAN *COUNTRY RANK* MENGUNAKAN ALGORITMA C4.5**

Erdwika Putri

Program Studi Statistika Fakultas MIPA

Universitas Islam Indonesia

## **INTISARI**

Dewasa ini, salah satu cara untuk mengkomunikasikan hasil-hasil karya ilmiah secara luas adalah dengan melalui tulisan, baik yang berupa suatu laporan ilmiah maupun tulisan-tulisan dalam majalah ilmiah. Menurut pasal 8 Permenpan No 17/2013, publikasi ilmiah dapat berbentuk buku referensi, buku ajar, monograf, artikel di media massa, serta jurnal ilmiah nasional dan internasional. Di Indonesia, publikasi ilmiah internasional lebih rendah dibandingkan dengan negara-negara maju dan negara-negara tetangga, padahal dengan melakukan publikasi ilmiah internasional dapat meningkatkan reputasi suatu negara. Salah satu entitas publikasi ilmiah internasional yang paling dikenal oleh para peneliti dunia adalah *Scopus*, dimiliki oleh *Elsevier* yang mempunyai dua indikator/variabel utama penilaian kualitas jurnal internasional, salah satunya yaitu *Scimago Journal & Country Rank (SJR)*. Penelitian ini bertujuan untuk melihat gambaran umum jurnal yang telah terindeks dalam *SJR*, mendeskripsikan hasil klasifikasi kategori *quartile* dan hasil akurasi klasifikasi tersebut, dengan metode *Decision Tree* algoritma C4.5. Dari analisis klasifikasi menggunakan algoritma C4.5 didapatkan gambaran umum bahwa jurnal dari Negara Indonesia masih sedikit yang terindeks oleh *SJR*, hasil klasifikasi yang berupa pohon keputusan dengan 66 terminal *node*, dan akurasi klasifikasi yang didapatkan dengan *10-fold cross validation* sebesar 70,97 %.

**Kata Kunci** : Algoritma C4.5, *Decision Tree*, Jurnal Ilmiah, Klasifikasi, *Scopus*, *SJR*.

**THE DATA CLASSIFICATION OF SCIMAGO JOURNAL AND COUNTRY  
RANK BY USING C4.5 ALGORITHM**

*Erdwika Putri*

*Statistics Study Program at Faculty of Mathematics and Natural Sciences  
Islamic University of Indonesia*

**ABSTRACT**

*Nowadays, one way to communicate the results of scientific work widely is through writing, either in the form of a scientific report or writings in scientific magazines. According to article 8 of Regulation of The Minister of The State Apparatus Utilization (Permenpan) No 17/2013, scientific publications can be in the form of reference books, textbooks, monographs, articles in the mass media, as well as national and international scientific journals. In Indonesia, the international scientific publications are lower than those of developed countries and neighboring countries, where as doing international scientific publications can improve the reputation of a country. One of the most familiar entities by the world researchers is Scopus owned by Elsevier, one of the world's leading publishers of scientific journals, which has two indicators / main variables to assess the quality of an international journal - one of which is Scimago Journal & Country Rank (SJR). This study, however, analyzes the overview of journals that have been indexed in the SJR, describes the results of the classification of the quartile category and the results of the classification accuracy, by using the Decision Tree method of C4.5 algorithm. From the classification analysis using the C4.5 algorithm, this research concludes that the Indonesia is the country still slightly indexed by the SJR; the classification result which is in the form of decision tree with 66 terminal nodes; lastly, the classification accuracy with 10-fold-cross-validation which is obtained is 70, 97%.*

*Keywords: C4.5 Algorithm, Decision Tree, Scientific Journal, Classification, Scopus, SJR.*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dewasa ini, salah satu cara untuk mengkomunikasikan hasil-hasil karya ilmiah secara luas adalah dengan melalui tulisan, baik yang berupa suatu laporan ilmiah maupun tulisan-tulisan dalam majalah ilmiah. Membuat tulisan ilmiah berarti menulis berdasarkan mengarah ke pandangan yang selaras dan telah terbukti kebenarannya, sehingga bisa dipertanggung-jawabkan (Suryoputro, Rudi, & Sya'ban, 2012). Mempublikasikan jurnal ilmiah yang telah dihasilkan secara internasional adalah perlu bagi setiap peneliti. Di Indonesia, mempublikasikan hasil dari karya ilmiah sendiri sangatlah penting dan merupakan salah satu aktivitas yang *inherent* dan melekat pada status peneliti, yang berarti berkat status itulah yang akan membuat seseorang harus melakukan publikasi ilmiah. Hanya melalui karya ilmiah yang dihasilkan, maka seorang peneliti dapat membangun nilai akademik yang baik, seperti terbiasa melakukan *critical thinking*, melakukan riset dengan baik, memiliki kemampuan analisis dan solusi, dan memiliki kemampuan beragumen, serta dapat menyalurkan pemikiran dengan baik serta dapat mengasai teknik menulis ilmiah (Kriyantono, 2015). Menurut pasal 8 Permenpan No 17/2013, publikasi ilmiah dapat berbentuk buku referensi, buku ajar, monograf, artikel di media massa, serta jurnal ilmiah nasional dan internasional (Permenpan, 2013).

Pertumbuhan akan cakupan dari penelitian telah merambah ke ranah yang lebih bersifat global, inerdisciplin, dan kolaboratif, sehingga memerlukan sebuah pusat data (*database*) yang akan mencakup fenomena ilmiah tersebut. Pusat data tersebut memiliki kegunaan untu dapat melihat apa yang sudah diteliti oleh peneliti lain dan mengetahui kontribusi apa yang harus diberikandari penelitian yang sedang dikerjakan, atau dengan kata lain untuk melihat sejauh mana kontribusi ilmiah yang

bisa ditawarkan kepada jurnal agar hasil karya peneliti dapat dipublikasikan (Elsevier, 2018).

Salah satu entitas yang paling dikenal oleh para peneliti dunia adalah *Scopus* yang dimiliki oleh *Elsevier*, salah satu penerbit utama jurnal ilmiah dunia. *Scopus* sendiri merupakan sebuah pusat data terbesar di dunia yang mencakup puluhan juta literatur ilmiah yang telah menerbitkan jurnal ilmiah sejak puluhan tahun yang lalu sampai sekarang. Fungsi utama dari *Scopus* adalah membuat indeks literatur ilmiah untuk memberikan informasi yang akurat mengenai *metadata* masing-masing artikel ilmiah secara individual, yang termasuk di dalamnya adalah data publikasi, abstrak, referensi, dan lainnya. Selain itu, *scopus* juga memberikan data yang agregat untuk memperlihatkan tingkat pengaruh suatu jurnal (*jurnal impact*) atau institusi (*institutional impact*) dalam dunia publikasi ilmiah dari dan ke artikel-artikel yang diterbitkan oleh sebuah jurnal atau dipublikasikan oleh peneliti-peneliti dari suatu institusi, sehingga pengguna *Scopus* dengan mudah bisa mendapatkan informasi mengenai apa yang sudah dipublikasikan oleh penerbit-penerbit atau lembaga-lembaga riset dari seluruh dunia (Elsevier, 2018).

*Scopus* adalah lembaga pengindeks yang mengevaluasi unjuk kerja (kualitas) sebuah jurnal internasional. Di Indonesia, sebuah jurnal dikatakan sebagai jurnal internasional bereputasi “terakreditasi” salah satunya bila terindeks di *Scopus*. SCOPUS memiliki/menciptakan 2 (dua) indikator/variabel utama penilaian untuk kerja sebuah jurnal internasional:

1. **SNIP** (*Source Normalized Impact per-Paper*) → Didefinisikan sebagai perbandingan jumlah situasi jurnal per-makalah yang diterbitkan sebuah jurnal dan potensi sitasi pada bidang ilmu yang sama (ranking dinormalisasikan pada bidang ilmu yang sama). Hal ini bertujuan untuk memungkinkan perbandingan/perankingan terhadap seluruh jurnal pada seluruh bidang keilmuan yang berbeda.
2. **SJR**: *Scimago Journal & Country Rank* (SJR) → Ukuran pengaruh ilmiah (*scientific influence*) sebuah jurnal yang mempertimbangkan 2 hal: (i) JUMLAH artikel yang merujuk, dan (ii) prestise/POPULERITAS dari

jurnal lain yang merujuk jurnal kita. *Scimago* adalah sebuah kelompok riset di Universitas Granada (Spanyol) yang merupakan partner *Scopus*, yang berdedikasi untuk analisis informasi, representasi dan perbaikan teknik visualisasi.

Di Indonesia, publikasi ilmiah internasional lebih rendah dibandingkan dengan negara-negara maju dan negara-negara tetangga. Hal ini disebabkan oleh minimnya akses referensi untuk penulisan publikasi, yang mana dengan adanya berbagai referensi tersebut, maka dapat memperkuat *state of the art* perkembangan sebuah penelitian; sarana dan prasarana penelitian yang kurang mendukung; kolaborasi penelitian yang masih sedikit, kolaborasi merupakan kunci utama untuk peningkatan kualitas publikasi; pembimbingan publikasi internasional, yang mana kriteria jurnal dan aturan penulisan di jurnal internasional saat ini masih belum dipahami oleh beberapa peneliti di Indonesia, sehingga banyak tulisan yang baik ada di jurnal yang dalam dan kualitas tulisan juga belum sesuai dengan kaidah yang dipersyaratkan; minimnya jurnal Indonesia bereputasi internasional; serta biaya dan intensif publikasi di jurnal internasional yang masih kurang. Biaya publikasi yang tinggi merupakan salah satu permasalahan yang sering dikeluhkan oleh peneliti ketika ingin menerbitkan artikel di jurnal internasional bereputasi. Selain itu, lembaga juga belum banyak yang memberikan penghargaan berupa insentif kepada peneliti yang publikasinya sudah diterbitkan di jurnal internasional bereputasi.

Berdasarkan uraian permasalahan mengenai publikasi ilmiah di jurnal internasional diatas, maka dalam penelitian ini penulis akan melakukan analisis dengan judul “Klasifikasi Data *Scimago Journal* Dan *Country Rank* Menggunakan Algoritma C4.5”. Penggunaan metode klasifikasi *decision tree* karena dapat mempermudah dalam pengambilan keputusan masalah yang kompleks menjadi lebih sederhana. Metode ini merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang diskrit, dan dibangun menggunakan algoritma C4.5.

Algoritma C4.5 ini merupakan teknik klasifikasi yang dapat membentuk pohon keputusan dengan tingkat keakuratan, kecepatan dan kehandalan yang lebih

baik dibanding teknik klasifikasi lainnya seperti *support vector machine* (SVM), *naive bayes*, dan *neural network* (Soewono dkk, 2014).

## 1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang diatas, maka didapatkan rumusan masalah sebagai berikut :

- a. Bagaimana gambaran umum jurnal yang telah terindeks ke dalam *Scimago Journal & Country Rank* ?
- b. Bagaimana hasil klasifikasi menggunakan metode algoritma C4.5 untuk kategori *quartile* pada jurnal SJR tersebut ?
- c. Bagaimana hasil akurasi klasifikasi dari metode algoritma C4.5 untuk kategori *quartile* pada jurnal SJR ?

## 1.3 Batasan Penelitian

Adapun batasan-batasan masalah pada penelitian ini adalah sebagai berikut:

- a. Data yang digunakan dalam penelitian ini adalah data sekunder *ranking journal* pada *website scimagojr.com* pada tahun 2016.
- b. Data variabel yang digunakan dalam analisis hanya 12 variabel dan yang bertipe *journal and book series*.
- c. Program yang digunakan dalam membuat pohon klasifikasi dengan menggunakan *software R 3.4.4*, untuk analisis deskriptif dengan menggunakan *Microsoft Excel 2013*, sedangkan untuk membuat *map* persebaran negara menggunakan *Tableau Public 10.3*.

## 1.4 Tujuan Penelitian

Dari rumusan masalah yang telah diuraikan diatas, maka dapat diketahui tujuan dari penelitian ini yaitu :

- a. Untuk melihat bagaimana gambaran umum jurnal yang telah terindeks ke dalam *Scimago Journal & Country Rank*.
- b. Untuk mengetahui bagaimana hasil klasifikasi menggunakan metode algoritma C4.5 untuk kategori *quartile* pada jurnal SJR tersebut.
- c. Untuk mengetahui bagaimana hasil akurasi klasifikasi dari metode algoritma C4.5 untuk kategori *quartile* pada jurnal SJR ?

### **1.5 Manfaat Penelitian**

Adapun dari penelitian yang telah dilakukan ini diharapkan dapat menghasilkan manfaat sebagai berikut :

1. Memberikan gambaran umum mengenai jurnal dan *book series* Indonesia yang ada di SJR.
2. Dapat dijadikan referensi bagi ristekdikti untuk mendukung para peneliti Indonesia yang ingin menerbitkan penelitiannya ke ranah internasional.

## **BAB II**

### **TINJAUAN PUSTAKA**

Klasifikasi merupakan pengelompokan secara sistematis pada suatu objek atau benda ke dalam golongan atau pola-pola tertentu berdasarkan kesamaan ciri. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari, baik pada bidang industri, sosial, kesehatan, maupun pendidikan. Penyelesaian masalah klasifikasi dapat dilakukan dengan metode klasifikasi. Klasifikasi menggunakan metode *Decision Tree* dengan Algoritma C4.5 telah banyak dilakukan dalam berbagai penelitian sebelumnya, namun klasifikasi ranking jurnal belum banyak dilakukan. Oleh karena itu, pada penelitian ini penulis melakukan *review* literatur yang mendukung penelitian ini.

Terdapat beberapa penelitian yang pernah dilakukan mengenai klasifikasi pada suatu objek atau benda tertentu diantaranya adalah penelitian yang dilakukan oleh Indranandita, dkk (2008) tentang klasifikasi dan pencarian terhadap jurnal dengan menggunakan metode *Vector Space Model* dan *Naïve Bayes*. Dari hasil penelitian yang telah dilakukan menggunakan kedua metode tersebut, disebutkan bahwa metode *Naïve Bayes* dengan FS-4 menghasilkan *precision* sebesar 64%, sedangkan metode *VSM* dengan pendekatan *Cosine* dalam melakukan pencarian menghasilkan *recall* sebesar 54,8% dan *precision* sebesar 60.7%, namun pada penelitian ini terdapat kekurangan yaitu perbaikan struktur data yang masih kurang, sehingga mengakibatkan proses *text mining* yang lama.

Oktaviana (2016) telah melakukan penelitian klasifikasi pola nasabah bank BRI Batang dengan menggunakan metode Algoritma C4.5. Penelitian tersebut bertujuan untuk menerapkan pengklasifikasi *data mining* pada data nasabah kredit pegawai BRI untuk mengetahui pola nasabah pada pembayaran kredit. Berdasarkan dari analisis yang telah dilakukan, hasil akurasi yang didapatkan dengan menggunakan algoritma C4.5 adalah sebesar 89,5%, yang mana

pada hasil pengamatan dari evaluasi ini diketahui bahwa kriteria macet yang diperoleh adalah yang berpenghasilan rendah atau yang berpenghasilan sedang dan dengan jangka waktu yang  $\leq 140$  dan atau yang jangka waktunya  $\leq 240$  dan berumur sudah tua. Kekurangan pada penelitian ini yaitu jumlah data *record* yang terbatas, sehingga *rules* yang terbentuk belum terlalu baik dan dapat meningkatkan hasil akurasi yang lebih tinggi.

Pada penelitian yang dilakukan oleh Sunjana (2010) tentang klasifikasi IPK mahasiswa dari beberapa mata kuliah dengan menggunakan *decision tree*. Hasil akurasi dari berbagai mata kuliah yang diuji dan didapatkan dengan menggunakan *decision tree* dalam penelitian ini diatas 50% semua, yang artinya bahwa pada kasus ini dengan menggunakan *decision tree* nilai IPK seorang mahasiswa terlihat sangat terpengaruh dengan 9 mata kuliah yang dianggap pokok, namun penelitian ini memiliki kekurangan yaitu data yang digunakan masih sedikit, sehingga hasil akurasi yang didapatkan masih kecil, oleh karena itu diharapkan untuk penelitian selanjutnya dapat menggunakan data yang lebih besar untuk meningkatkan akurasi serta presentase kebenaran pada *tree* yang dibuat.

Mashlahah (2013) telah melakukan penelitian klasifikasi kelulusan mahasiswa dengan menggunakan metode klasifikasi yaitu algoritma C4.5. Penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa. Hasil akurasi yang telah didapatkan dalam penelitian ini yaitu dengan data sampel 60, akurasinya sebesar 64,51%, pada sampel 79, akurasinya sebesar 70,96%, dan ketika sampelnya sebanyak 90, akurasinya sebesar 82,97%. Kekurangan pada penelitian ini yaitu data yang digunakan masih sedikit.

Harryanto dan Hansun (2017) telah melakukan penelitian klasifikasi penerimaan calon pegawai baru di PT WISE dengan menerapkan algoritma C4.5. Penelitian ini bertujuan untuk memprediksi proses penerimaan calon pegawai baru pada PT WISE. Hasil tingkat keberhasilan prediksi calon pegawai baru di PT WISE secara keseluruhan yang telah diukur menggunakan metode *ten-fold cross validation* adalah sebesar 71%, namun pada penelitian ini juga terdapat kekurangan

salah satunya yaitu masih sedikitnya sampel data yang digunakan, sehingga tingkat akurasi yang dihasilkan masih belum tinggi.

Supriyanti, Kusriani, dan Amborowati (2016) dengan penelitiannya mengenai perbandingan kinerja algoritma C4.5 dan Naive Bayes untuk ketepatan pemilihan konsentrasi mahasiswa. Pada penelitian tersebut diperoleh hasil kinerja antara algoritma C4.5 tanpa penambahan seleksi fitur *forward selection* dengan algoritma C4.5 ditambah seleksi fitur *forward selection* lebih unggul bila dibandingkan dengan algoritma Naive Bayes pada kasus ketepatan pemilihan konsentrasi mahasiswa, akurasi untuk algoritma C4.5 tanpa fitur *forward selection* sebesar 84,43% dan setelah ditambahkan fitur *forward selection* meningkat menjadi 84,98%, sedangkan untuk Naive Bayes dari 78,47% meningkat menjadi 82,01%. Kekurangan pada penelitian ini yaitu penggunaan variabel yang masih sedikit.

Berdasarkan uraian diatas, maka dapat dilihat perbandingan dari beberapa penelitian sebelumnya yang akan digunakan sebagai tinjauan pustaka dalam penelitian ini. Untuk mempermudah dalam membandingkan penelitian-penelitian yang telah dijabarkan diatas, maka dibuatlah tabel 2.1 berikut.

**Tabel 2.1** Perbandingan Penelitian

No	Penelitian	Metode	Hasil
1	Indranandita, Susanto, dan Rachmat (2008)	Klasifikasi dengan <i>Vector Space Model</i> dan <i>Naive Bayes</i>	Hasil didapatkan bahwa metode <i>Naive Bayes</i> dengan FS-4 menghasilkan <i>precision</i> sebesar 64%, sedangkan metode <i>VSM</i> dengan pendekatan <i>Cosine</i> dalam melakukan pencarian menghasilkan <i>recall</i> sebesar 54,8% dan <i>precision</i> sebesar 60,7%, namun pada penelitian ini terdapat kekurangan yaitu

No	Penelitian	Metode	Hasil
			perbaikan struktur data yang masih kurang, sehingga mengakibatkan proses <i>text mining</i> yang lama.
2	Oktaviana (2016)	Klasifikasi dengan Algoritma C4.5	Hasil akurasi yang didapatkan dengan menggunakan algoritma C4.5 adalah sebesar 89,5%, yang mana pada hasil pengamatan dari evaluasi ini diketahui bahwa kriteria macet yang diperoleh adalah yang berpenghasilan rendah atau yang berpenghasilan sedang dan dengan jangka waktu yang $\leq 140$ dan atau yang jangka waktunya $\leq 240$ dan berumur sudah tua. Kekurangan pada penelitian ini yaitu jumlah data <i>record</i> yang terbatas, sehingga <i>rules</i> yang terbentuk belum terlalu baik dan dapat meningkatkan hasil akurasi yang lebih tinggi.
3	Sunjana (2010)	Klasifikasi dengan <i>decision tree</i>	Didapatkan hasil akurasi dalam penelitian ini diatas 50% semua, yang artinya bahwa pada kasus ini dengan menggunakan <i>decision tree</i> nilai IPK seorang mahasiswa

No	Penelitian	Metode	Hasil
			terlihat sangat terpengaruh dengan 9 mata kuliah yang dianggap pokok, namun penelitian ini memiliki kekurangan yaitu data yang digunakan masih sedikit, sehingga hasil akurasi yang didapatkan masih kecil, oleh karena itu diharapkan untuk penelitian selanjutnya dapat menggunakan data yang lebih besar untuk meningkatkan akurasi serta presentase kebenaran pada <i>tree</i> yang dibuat.
4	Mashlahah (2013)	Klasifikasi dengan algoritma C4.5.	Didapatkan hasil, jika data sampel 60, akurasinya sebesar 64,51%, pada sampel 79, akurasinya sebesar 70,96%, dan ketika sampelnya sebanyak 90, akurasinya sebesar 82,97%. Kekurangan pada penelitian ini yaitu data yang digunakan masih sedikit.
5	Harryanto dan Hansun (2017)	Klasifikasi Algoritma C4.5	Hasil tingkat keberhasilan prediksi calon pegawai baru di PT WISE secara keseluruhan yang telah diukur menggunakan metode <i>ten-fold</i>

No	Penelitian	Metode	Hasil
			<p><i>cross validation</i> adalah sebesar 71%, namun pada penelitian ini juga terdapat kekurangan salah satunya yaitu masih sedikitnya sampel data yang digunakan, sehingga tingkat akurasi yang dihasilkan masih belum tinggi.</p>
6	Supriyanti, Kusrini, dan Amborowati (2016)	Klasifikasi Algoritma C4.5 dan Naive Bayes	<p>Kinerja antara algoritma C4.5 tanpa penambahan seleksi fitur <i>forward selection</i> dengan algoritma C4.5 ditambah seleksi fitur <i>forward selection</i> lebih unggul bila dibandingkan dengan algoritma Naive Bayes pada kasus ketepatan pemilihan konsentrasi mahasiswa, akurasi untuk algoritma C4.5 tanpa fitur <i>forward selection</i> sebesar 84,43% dan setelah ditambahkan fitur <i>forward selection</i> meningkat menjadi 84,98%, sedangkan untuk Naive Bayes dari 78,47% meningkat menjadi 82,01%. Kekurangan pada penelitian</p>

No	Penelitian	Metode	Hasil
			ini yaitu penggunaan variabel yang masih sedikit.

Berdasarkan uraian dari tinjauan pustaka diatas, maka penelitian yang akan penulis lakukan yaitu pengklasifikasian jurnal internasional yang telah terindeks *Scopus* dalam *Scimago Journal & Country Rank* yang berjudul “Klasifikasi Data *Scimago Journal & Country Rank* Menggunakan *Decision Tree* Dengan Algoritma C4.5”. Metode yang digunakan dalam pengklasifikasian ini yaitu *decision tree* dengan algoritma C4.5, yang mana penggunaan metode tersebut karena hasil pengklasifikasian dengan pohon keputusan dapat dipahami dengan mudah oleh masyarakat awam dan dengan menggunakan algoritma C4.5 dikarenakan algoritma tersebut dapat mengatasi data yang berupa numerik ataupun kategorik.

## **BAB III**

### **LANDASAN TEORI**

#### **3.1 Jurnal Ilmiah**

Jurnal ilmiah adalah salah satu bentuk publikasi ilmiah yang diterbitkan secara berkala yang memuat hasil kegiatan dalam bidang keilmuan tertentu, baik berupa hasil pengamatan secara empirik maupun kajian konseptual yang bersifat penemuan baru, maupun koreksi, pengembangan, atau penguatan terhadap paradigma, konsep, prinsip, hukum, dan teori yang telah ada. Adnan, dkk (2005) menyebutkan bahwa jurnal ilmiah itu sebagai forum komunikasi untuk anggota masyarakat dengan disiplin ilmu tertentu, karena jurnal tersebut dibaca oleh anggota masyarakat tertentu, maka jurnal ilmiah harus menyiapkan artikel-artikel yang sesuai dengan minat dan kepentingan tersebut (Adnan & Zifirdaus, 2005).

Isi dari jurnal ilmiah berupa artikel ilmiah (*research articles*) yaitu tulisan yang berisi laporan yang disusun secara sistematis mengenai hasil dari kajian atau hasil dari penelitian yang disajikan bagi masyarakat ilmiah tertentu, yang merupakan audiens khusus yang bertujuan menyampaikan hasil kajian dan kontribusi penulis artikel kepada pembaca untuk dipikirkan, dikaji kembali, dan diperdebatkan, baik secara lisan maupun secara tertulis. Yang dimaksud dengan hasil kajian adalah hasil pemikiran intensif tentang suatu topik, sedangkan hasil penelitian umumnya lebih spesifik karena harus melibatkan data yang dipublikasikan di jurnal ilmiah, laporan dari surat kabar, atau majalah, wawancara, laporan saksi mata, dokumen dan sebagainya (Adnan & Zifirdaus, 2005).

Jurnal ilmiah diharapkan ditulis berdasarkan fakta dan dapat memberikan manfaat, meskipun penulisan jurnal tersebut sudah dilakukan dalam waktu yang cukup lama. Dengan kata lain, tetap bisa dimanfaatkan sebagai sumber pengetahuan bagi yang pembacanya. Artikel hasil penelitian merupakan tulisan yang paling

sering dimuat dalam jurnal ilmiah, sehingga ada identifikasi bahwa jurnal ilmiah adalah kumpulan artikel hasil penelitian ilmiah.

Jurnal ilmiah dapat terpecah menjadi jurnal ilmiah nasional dan jurnal ilmiah Internasional. Di Indonesia misalnya, menurut pedoman Operasional Angka Kredit Dikti 2014 jurnal ilmiah nasional itu sendiri adalah sebuah jurnal yang memiliki (1) ISSN; (2) Menyebarkan hasil penelitian atau konsep ilmiah; (3) Ditujukan pada masyarakat ilmiah/peneliti; (4) Diterbitkan oleh badan ilmiah/organisasi/perguruan tinggi; (5) Mempunyai dewan redaksi yang ahli dalam bidangnya dan telah melalui *peer review process*; (6) Ditulis menggunakan Bahasa Indonesia & atau Bahasa Inggris dengan abstraknya Bahasa Indonesia; dan (8) Diedarkan nasional. Sementara itu, jurnal internasional dibedakan menjadi dua, yaitu jurnal internasional dan jurnal internasional bereputasi. Jurnal internasional adalah jurnal yang memiliki ciri-ciri (1) Karya yang diterbitkan berdasarkan tulisan yang sudah memenuhi kaidah ilmiah dan etika keilmuan; (2) Memiliki ISSN dan telah beredar di berbagai negara; (3) Tulisannya menggunakan bahasa resmi PBB (Arab, Inggris, Perancis, Rusia, Spanyol dan Tiongkok); (4) Mempunyai terbitan dalam versi *online*. Terdapat dua jenis jurnal internasional, yang pertama yaitu jurnal tersebut telah terindeks *Web of Science*, *Scopus*, atau *Microsoft Academic Search* tapi belum mempunyai faktor dampak (*impact factor*) dari *ISI Web of Science (Thomson Reuters)* atau *Scimago Journal & Country Rank (SJR)*. Jurnal internasional yang kedua adalah yang belum terindeks *Web of Science*, *Scopus*, atau *Microsoft Academic Search* tetapi, telah terindeks pada *database* internasional seperti *DOAJ*, *CABI*, *Copernicus*, atau laman yang sesuai dengan pertimbangan dari Ditjen Dikti *DOAJ*, *Copernicus*, *Ebscho* (Kriyantono, 2015).

Jurnal internasional bereputasi yaitu jurnal yang telah memenuhi kriteria jurnal internasional dan dengan kriteria tambahan (1) Telah terindeks pada *Web of Science* atau *Scopus*; (2) Mempunyai faktor dampak (*impact factor*) dari *ISI Web of Science (Thomson Reuters)*, *Scimago Journal & Country Rank (SJR)*, *Art & Humanities Citation Index (AHCI)*.

### 3.2 Scimago Journal & Country Rank

*Scopus* adalah sebuah *database indexing*, yang berisi *abstract* dan *citations* publikasi yang telah banyak dipercaya sebagai salah satu tolak ukur, yang dikelola oleh *Elsevier*. *Scopus* mengeluarkan *Scimago Journal & Country Rank* (SJR). *Scimago Journal & Country Rank* adalah ukuran pengaruh ilmiah (*scientific influence*) yang mempertimbangkan dalam dua hal, yaitu (1) Jumlah artikel yang merujuk dan; (2) *prestise*/popularitas dari jurnal lain yang merujuk jurnal kita. Dalam penentuan ranking pada SJR didasarkan dari nilai Q, yang mana Q tersebut merupakan kuartil untuk setiap jurnal pada masing-masing kategori subjek yang ada. Q1 adalah *cluster* paling tinggi atau paling utama dari sisi kualitas jurnal, kemudian diikuti Q2, Q3, dan Q4 (Scimagojr, 2018). Penentuan nilai kuartil sendiri didasarkan pada persentil dari masing-masing kategori subjek jurnal, yang mana kategorinya yaitu :

$$\text{Quartile 1} = 75^{\text{th}} - 99^{\text{th}}$$

$$\text{Quartile 2} = 50^{\text{th}} - 74^{\text{th}}$$

$$\text{Quartile 3} = 25^{\text{th}} - 49^{\text{th}}$$

$$\text{Quartile 4} = 0 - 24^{\text{th}}$$

Kategori subjek jurnal pada SJR berdasarkan kategori yang ada pada *Scopus*, yang mana pada *Scopus* tersebut terdapat 4 kategori, diantaranya yaitu (1) *Social Science*; (2) *Health Science*; (3) *Physical Science*; dan (4) *Life Science* (Elsevier, 2018). Dari keempat kategori tersebut dibagi lagi menjadi beberapa sub-subjek yang nantinya akan mengelompokkan judul-judul dari jurnal yang ada. Tabel 3.1 berikut merupakan tampilan dari kategori subjek dan sub-subjek pada SJR :

**Tabel 3.1** Kategori Subjek Pada SJR

Subjek	Sub-Subjek
Social Science	Arts & Humanities
	Business, Management & Accounting
	Economics, Econometrics & Finance
	Psychology
	Multidisciplinary

Subjek	Sub-Subjek
	Social Science
Health Science	Medicine
	Nursing
	Dentistry
	Health Professions
	Veterinary
Physical Science	Chemical Engineering
	Chemistry
	Computer Science
	Decision Sciences
	Earth & Planetary Sciences
	Mathematics
	Physics & Astronomy
Life Science	Agricultural & Biological Science
	Biochemistry, Genetics & Molecular Biology
	Neuroscience
	Immunology & Mikrobiology
	Pharmacology, Toxicology & Pharmaceutics
	Environmental Science

### 3.3 Data Mining

*Data mining* adalah istilah yang biasa dipakai untuk memaparkan penemuan pengetahuan di dalam *database*. *Data mining* adalah suatu proses yang memakai teknik statistik, matematika, kecerdasan tiruan, dan *machine learning* untuk mengekstraksi serta mengidentifikasi informasi yang berguna untuk pengetahuan yang terkait dari berbagai *database* besar (Turban, 2015). *Data mining* juga merupakan suatu proses untuk menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa sekumpulan data besar/banyak yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematik (Larose & Larose, 2005).

Teknik dalam *data mining* dapat memungkinkan untuk menciptakan sebuah model untuk mengenali pola data yang lain selain dalam basis data yang tersimpan. Selain itu, *data mining* juga bisa dimanfaatkan untuk memprediksi ataupun klasifikasi yang bertujuan agar terbentuk pola universal dari data-data yang ada (Haryati, Sudarsono, & Suryana, 2015).

Ada 3 teknik yang sering digunakan dalam *data mining* (Haryati, Sudarsono, & Suryana, 2015) :

1. *Association Rule Mining*

*Association rule mining* merupakan salah satu teknik dalam *data mining* yang digunakan untuk menentukan hubungan antara kombinasi atribut. Contohnya adalah analisa pembelian di swalayan yang akan menghasilkan strategi dalam penempatan barang di swalayan tersebut atau dapat menggunakan diskon untuk kombinasi dari pembelian barang tertentu.

2. *Clustering*

*Clustering* biasanya digunakan untuk memberikan label pada kelas data yang belum diketahui, sehingga sering disebut dengan *unsupervised learning*. Prinsip yang sering digunakan dalam *clustering* adalah dengan memaksimalkan tingkat kesamaan antar *cluster*.

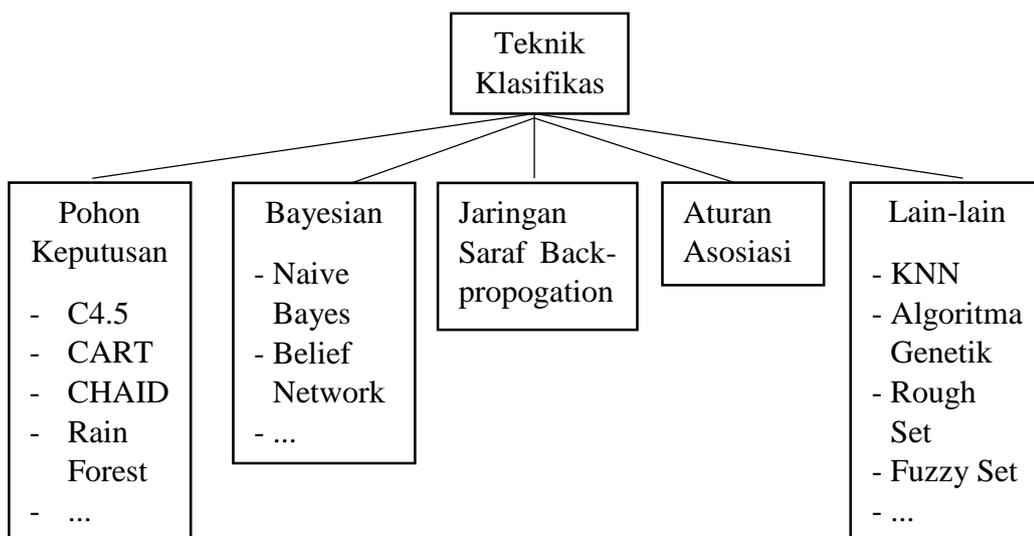
3. *Classification*

Hasil dalam klasifikasi adalah adanya variabel kategori.

### 3.4 Klasifikasi

Klasifikasi adalah suatu proses yang digunakan untuk menemukan sebuah model atau fungsi yang dapat menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memprediksi kelas dari suatu objek baru yang labelnya tidak diketahui. Model yang dimaksud dapat berupa aturan “jika-maka”, *decision tree*, formula matematis atau *neural network*. Proses klasifikasi sendiri biasanya terbagi menjadi dua fase, yaitu *training* dan *testing*. Fase *training* berarti, sebagian dari data yang ada dan telah diketahui kelas datanya dilatih untuk membentuk sebuah model prediksi. Pada fase *testing*, model yang telah didapatkan sebelumnya diuji dengan sebagian data lainnya untuk mengetahui keakuratan dari model tersebut. Bila akurasi yang didapatkan sudah mencukupi, maka model yang terbentuk tersebut dapat dipakai untuk prediksi kelas data yang belum diketahui. Terdapat empat komponen dasar dalam proses klasifikasi (Gorunescu, 2011) :

1. Kelas (*class*) atau label kelas adalah variabel dependen dari model yang merupakan variabel kategori dan menjelaskan sebuah “label” pada objek setelah klasifikasi. Contoh : SJR Best Quartile (label : Q1, Q2, Q3, dan Q4), kelas bintang (galaksi), dan lainnya.
2. Preiktor (*predictor*) atau atribut (*attribute*) merupakan variabel independen dari model yang diwakili oleh karakteristik (atribut) dari data yang akan diklasifikasikan. Contoh : SJR, H index, Total Ref, dan lain-lain.
3. *Training set* atau dataset latih merupakan sekumpulan data yang didalamnya berisi nilai atau *record* untuk dua komponen sebelumnya (kelas dan prediktor) yang dapat berupa variabel kontinyu maupun kategorik, dan sebagian dari sekumpulan data tersebut digunakan untuk “pelatihan” atau pembuatan model yang disesuaikan dengan kelasnya berdasarkan prediktor yang tersedia.
4. *Testing set* atau dataset pengujian yaitu berisi data baru yang akan diklasifikasikan oleh model yang telah didapatkan dan untuk mengukur tingkat keakuratan klasifikasi, sehingga performansi model klasifikasi dapat dievaluasi.



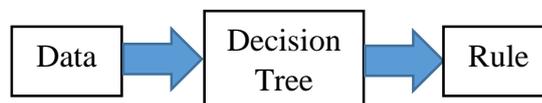
**Gambar 3.1** Teknik Klasifikasi (Soewono, Gemowo, & Sasongko, 2014)

### 3.5 Data Training dan Data Testing

Dalam klasifikasi data pada umumnya dibagi menjadi dua, yaitu *data training* dan *data testing*. Untuk membentuk sebuah model klasifikasi, maka dilakukan *data training* yang mana *data training* tersebut biasanya digunakan oleh algoritma klasifikasi (misalnya *decision tree*, bayesian, *neural network*, SVM). Representasi pengetahuan dari model yang telah dihasilkan tersebut dapat digunakan untuk mengukur sejauh mana tingkat keberhasilan dari klasifikasi tersebut melakukan klasifikasi dengan benar. Oleh karena itu, pada saat melakukan *testing*, data yang diuji seharusnya tidak terdapat pada data *training*, sehingga dapat diketahui apakah model klasifikasi dapat melakukan klasifikasinya dengan baik. Proporsi untuk melakukan data *training* biasanya lebih besar dibanding data *testing* dan biasanya 2/3 dari total data dijadikan sebagai data *training*, sedangkan sisanya akan digunakan sebagai data *testing*. Selain itu, ada juga penelitian yang menghasilkan tingkat keakuratan model klasifikasi optimum dengan menggunakan proporsi untuk data *training* dan *testing* sebesar 80:20 dan 90:10 (Paratu, 2013).

### 3.6 Pohon Keputusan (*Decision Tree*)

Pohon keputusan merupakan salah satu metode klasifikasi yang kuat dan terkenal. Metode pohon keputusan mengubah fakta yang besar menjadi pohon keputusan yang merepresentasikan aturan, dan aturan tersebut dapat dengan mudah untuk dipahami oleh manusia. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variabel *input* dengan sebuah variabel target (Berry & Linoff, 2004).

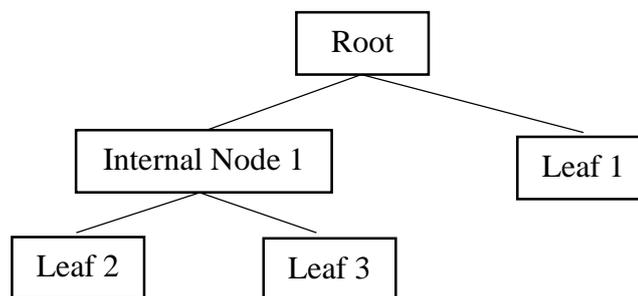


**Gambar 3.2** Konsep Pohon Keputusan

Konsep dari pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon keputusan, kemudian mengubah pohon menjadi bentuk *rule*, dan selanjutnya *rule* akan disederhanakan. Salah satu manfaat utama yang didapatkan dari penggunaan pohon keputusan adalah kemampuannya untuk mem-*break down*

proses pengambilan keputusan yang kompleks menjadi lebih simpel, sehingga pengambil keputusan akan lebih mudah dalam menginterpretasikan solusi dari permasalahan tersebut. Keuntungan lainnya dalam penggunaan pohon keputusan adalah pohon keputusan dapat memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal pemodelan (Haryati, Sudarsono, & Suryana, 2015).

Pohon (*tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*). Terdapat tiga simpul yang berbeda pada sebuah pohon, yaitu simpul akar (*root node*), simpul percabangan/internal (*branch/internal node*) dan simpul daun (*leaf node*) (Astuti, 2013). Dalam konsep dasar pohon keputusan, simpul akar (*root*) digunakan sebagai bagian awal dari pohon keputusan, sedangkan yang menjadi cabang dari pohon keputusan adalah *internal node*, dan titik akhir (*leaf*) merupakan pembagian kelas yang dihasilkan.



**Gambar 3.3** Konsep Dasar Pohon Keputusan

Dalam pohon keputusan banyak terdapat algoritma yang digunakan, dan seiring dengan berkembangnya zaman, maka algoritma pada pohon keputusan juga sudah banyak mengalami perkembangan dan sudah banyak digunakan dalam penelitian-penelitian sebelumnya, beberapa algoritma dalam pohon keputusan tersebut dapat dilihat pada tabel 3.2 berikut.

**Tabel 3.2** Frekuensi Penggunaan Algoritma Pohon Keputusan

Algoritma Pohon Keputusan	Frekuensi
ID3	68 %
C4.5	54.55 %
CART	40.9 %
SPRINT	31.84 %

Algoritma Pohon Keputusan	Frekuensi
SLIQ	27.27 %
PUBLIC	13.6 %
C5.0	9 %
CLS	9 %
RANDOM FOREST	9 %
RANDOM TREE	4.5 %
ID3+	4.5 %
OCI	4.5 %
CLOUDS	4.5 %

### 3.7 Algoritma C4.5

#### 3.7.1 Pengertian Algoritma C4.5

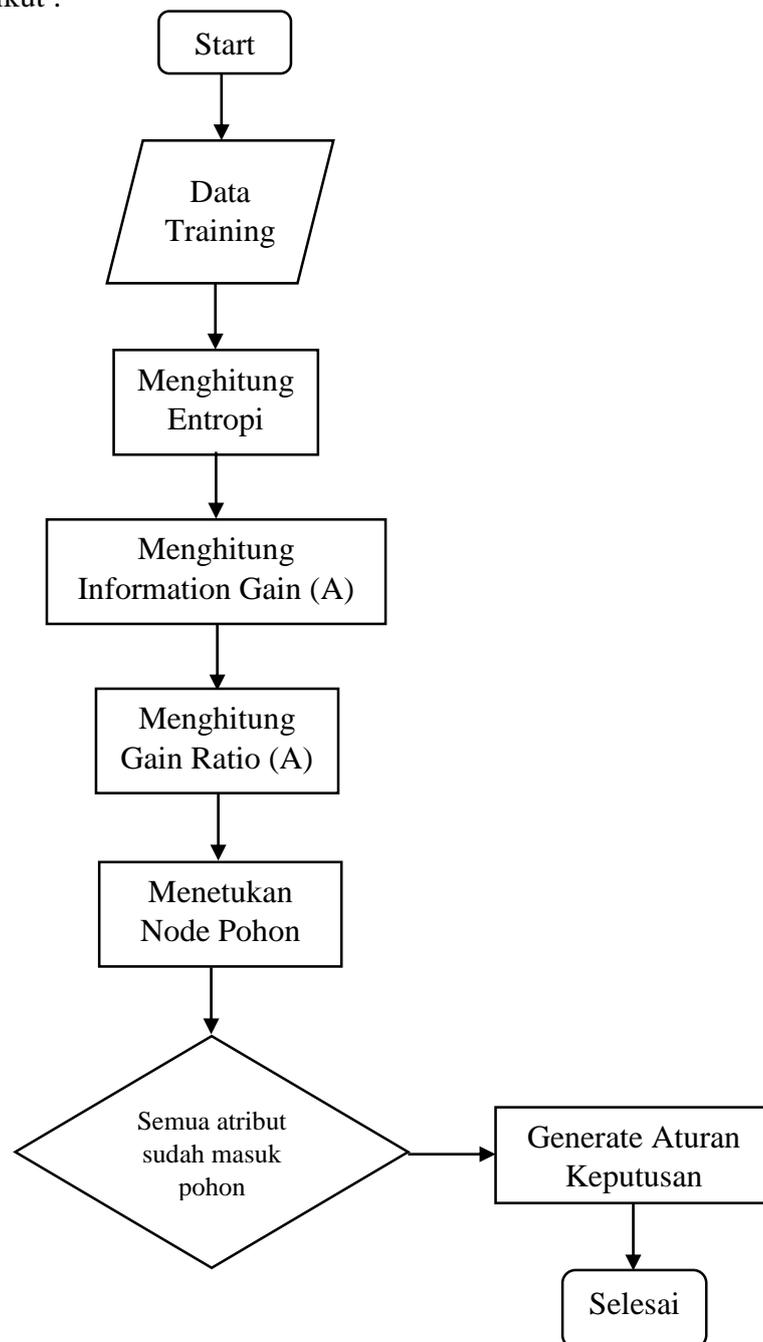
Algoritma C4.5 merupakan salah satu deretan algoritma yang digunakan untuk permasalahan klasifikasi pada sebuah mesin dan himpunan data. Dengan nilai-nilai yang beragam, dimana kejadian diuraikan oleh koleksi atribut dan memiliki salah satu dari satu set kelas yang eksklusif.

Algoritma C4.5 yang dikembangkan oleh J.Ross Quinlan (1983) ini merupakan generasi terbaru dari algoritma ID3. Dalam ID3, induksi dari *decision tree* hanya bisa dipakai pada fitur data yang bertipe kategorikal (nominal), sedangkan untuk tipe data numerik (rasio) tidak bisa digunakan. Perubahan yang membedakan antara algoritma C4.5 dan ID3 adalah algoritma C4.5 dapat menangani fitur data yang bertipe numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule set*. Dalam menentukan fitur yang akan menjadi pemecah *node* pada pohon yang diinduksi, algoritma C4.5 ini juga menggunakan dengan kriteria *gain*.

Algoritma C4.5 ini adalah salah satu teknik dari *decision tree* yang sering digunakan, karena dapat menghasilkan aturan-aturan dan sebuah pohon keputusan dengan tujuan untuk meningkatkan akurasi dari prediksi yang sedang dilakukan (Rani, 2015).

### 3.7.2 Langkah-langkah Algoritma C4.5

Dengan menggunakan algoritma C4.5 dalam pembuatan pohon keputusannya, ada beberapa langkah-langkah yang harus dilakukan, yaitu sebagai berikut :



**Gambar 3.4** Flowchart *Algoritma C4.5*

Berikut langkah-langkah dalam pembuatan pohon keputusan menggunakan algoritma C4.5 :

- Langkah 1** : Menentukan data *training*
- Langkah 2** : Menghitung nilai entropy pada seluruh data atau entropy total
- Langkah 3** : Menghitung nilai *information gain* pada masing-masing atribut
- Langkah 4** : Menghitung nilai *gain ratio* pada masing-masing atribut
- Langkah 5** : Memilih atribut yang akan digunakan sebagai *node* berdasarkan nilai *gain ratio* tertinggi
- Langkah 6** : Mengecek apakah semua atribut sudah terbentuk pada pohon, jika belum maka ulangi proses 2-5, jika sudah maka lanjut pada proses selanjutnya
- Langkah 7** : Aturan keputusan di-*generate* mengikuti pohon yang telah dibentuk sebelumnya.

### 3.7.3 Metode Pencarian Atribut Pada *Node*

Dalam membuat pohon keputusan menggunakan algoritma C4.5, terdapat beberapa komponen-komponen penyusunnya (Julianto, Yunitarini, & Sophan, 2014) :

a) *Entropy*

*Entropy* adalah suatu parameter yang digunakan untuk mengukur tingkat keberagaman (heterogenitas) dari sekumpulan data. Semakin heterogen suatu data, maka nilai entropi yang didapatkan akan semakin besar. Perhitungan nilai entropi dapat dilihat pada persamaan berikut ini :

$$Entropy (S) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (3.1)$$

Keterangan :

S : Himpunan Kasus

n : Banyaknya Kelas

$p_i$  : Proporsi untuk kelas ke-i

b) *Information Gain*

*Information Gain* biasa digunakan untuk memilih atribut uji pada setiap *node* didalam pohon. Atribut yang memiliki nilai *information gain* tertinggi akan dipilih sebagai *parent* untuk *node* selanjutnya. Secara matematis, *information gain* dari suatu atribut A dituliskan sebagai berikut :

$$Gain(A) = Entropy(S) - \sum_{j=1}^m p_j Entropy(S_j) \quad (3.2)$$

Keterangan :

A : Atribut

m : Banyaknya Kategorik

P<sub>j</sub> : Proporsi untuk kategorik j

Entropy (S<sub>j</sub>) : Entropy kategorik j pada setiap atribut

c) *Gain Ratio*

Dalam perhitungannya, *information gain* masih memiliki kekurangan, salah satunya yaitu kekurangan yang mungkin saja terjadi pada saat pemilihan atribut yang tidak relevan. Oleh karena itu, *gain ratio* yang merupakan normalisasi dari *information gain* akan memperhitungkan entropi dari distribusi probabilitas subset setelah dilakukannya proses partisi. Perhitungan dari *gain ratio* adalah sebagai berikut :

$$Gain Ratio(A) = \frac{Gain(A)}{Split Info(A)} \quad (3.3)$$

Yang mana *split info* adalah entropi dari seluruh distribusi probabilitas subset setelah dilakukannya pemartisian (*splitting*).

$$SplitInfo(A) = -\sum_{j=1}^n p_j \log_2(p_j) \quad (3.4)$$

### 3.7.4 Aturan Pemberhentian

Sebuah *node* akan berhenti mengalami proses pemisahan (*split*) dan akan menghasilkan daun, jika :

1. *Node* menjadi *pure*, yang artinya semua objek yang berada pada *node* tersebut mempunyai kelas yang sama.

2. Tidak ada atribut yang tersisa.

### 3.7.5 Confusion Matrix

Untuk mengevaluasi kinerja dari suatu metode atau model, maka diperlukannya sebuah cara yang sistematis. Pada evaluasi klasifikasi didasarkan pengujian pada objek yang benar dan salah (Gorunescu, 2011). Untuk menentukan jenis terbaik dari skema pembelajaran yang digunakan, maka menggunakan validasi data yang berdasarkan data pelatihan untuk melatih skema pembelajaran (Ian, Frank, & Mark, 2011).

*Confusion matrix* (Kohavi & Provost, 1998) berisi informasi mengenai hasil klasifikasi aktual dan yang telah diprediksi oleh sistem klasifikasi. Performa dari sistem tersebut biasanya dievaluasi menggunakan data dalam sebuah matriks. *Confusion matrix* juga merupakan tabel yang digunakan sebagai alat ukur yang berguna untuk melakukan analisis seberapa baik hasil pengklasifikasian yang benar dan salah dari hasil prediksi yang telah dilakukan dalam kelas yang berbeda-beda. Berikut adalah contoh *confusion matrix* untuk dua kelas :

**Tabel 3.3** *Confusion Matrix* 2 Kelas

Predicted Class	Actual Class	
	Col 1	Col 2
Row 1	Benar1	Salah1
Row 2	Salah2	Benar2

Pada Tabel 3.2 diatas adalah gambaran dari *confusion matrix*. Terdapat beberapa rumus umum yang bisa digunakan untuk menghitung performa klasifikasi. Hasil dari nilai akurasi, presisi, dan *recall* biasa ditampilkan dalam presentase (Rianto, 2016).

$$Akurasi = \frac{\sum \text{Prediksi benar}}{\sum (\text{benar} + \text{salah})} \times 100\% \quad (3.5)$$

$$Error Rate = \frac{\sum \text{Prediksi salah}}{\sum (\text{benar} + \text{salah})} \times 100\% \quad (3.6)$$

$$Presisi = \frac{Benar1}{Benar1 + Salah1} \times 100\% \quad (3.7)$$

$$Recall = \frac{Benar1}{Benar1+Salah2} \times 100\% \quad (3.8)$$

### 3.7.6 *K-fold Cross Validation*

*K-fold cross validation* merupakan salah satu teknik untuk melakukan estimasi tingkat kesalahan pengujian pemrosesan citra digital. Cara kerja *K-fold cross validation* yaitu dengan mengelompokkan data latih dan data uji yang saling terpisah, kemudian melakukan proses pengujian yang diulang sebanyak *K* kali (Fauzie, 2010).

Langkah dari *K-fold cross validation* antara lain : (1) Membagi data asli yang tersedia menjadi *K* kelompok; (2) Setiap *K* dibuat sejumlah *T* himpunan data yang memuat semua data latih kecuali yang berada di kelompok ke-*k*; (3) Mengerjakan algoritma yang dimiliki dengan sejumlah *T* data latih; (4) Pengujian algoritma menggunakan data pada kelompok *K* sebagai data uji; (5) Melakukan pencatatan hasil algoritma (Pratiwi, 2010). Keuntungan dari teknik *K-fold cross validation* ini yaitu menunjukkan bahwa semua elemen pada baris data digunakan untuk pelatihan sekaligus pengujian (Pratiwi, 2010).

## 3.8 Analisis Regresi

Analisis regresi dalam statistika adalah salah satu metode untuk menentukan hubungan sebab-akibat antara satu variabel dengan variabel-variabel yang lain. Analisis regresi adalah salah satu analisis yang paling populer dan luas pemakaiannya. Analisis regresi dipakai secara luas untuk melakukan prediksi dan ramalan. Analisis ini juga digunakan untuk memahami variabel bebas (independen) mana saja yang berhubungan dengan variabel terikat (dependen), serta untuk mengetahui bentuk-bentuk hubungan tersebut (Wikipedia, 2017).

Sebelum melakukan pengujian, terlebih dahulu kita harus mendefinisikan Hipotesis Awal ( $H_0$ ) dan Hipotesis Alternatif ( $H_1$ ), berikut merupakan langkah analisis regresi:

### Hipotesis

$H_0$  : Tidak ada pengaruh signifikan antara variabel independe terhadap variabel dependen yang diamati

$H_1$  : Ada pengaruh secara signifikan antara variabel independen terhadap variabel dependen yang diamati

## **BAB IV**

### **METODOLOGI PENELITIAN**

#### **4.1 Populasi Penelitian**

Populasi penelitian yang digunakan dalam penelitian adalah data *journal country ranking* tahun 2016 yang terdapat pada situs *website scimagojr* yang terdiri dari 16 variabel di dalamnya. Sampel yang digunakan dalam penelitian ini adalah semua data jurnal yang hanya masuk ke dalam kategori Q1, Q2, Q3, dan Q4 dan 11 variabel yang lain. Data jurnal pada Q1 berjumlah 6.757 data, pada Q2 berjumlah 6.032 data, untuk Q3 berjumlah 5.721 data, dan Q4 berjumlah 5.419 data, sehingga total semua data yang digunakan adalah sebanyak 23.929 data.

#### **4.2 Jenis dan Sumber Data**

Data yang digunakan pada penelitian ini berupa data sekunder yang didapatkan dari *website scimagojr.com*. Data sekunder adalah data yang diperoleh dari sumber yang ada, dalam hal ini sumbernya adalah dari *website scimagojr*.

#### **4.3 Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini dibagi menjadi dua, yaitu variabel dependen dan variabel independen. Pada Tabel 4.1 berikut akan ditampilkan tentang penjelasan dan definisi dari masing-masing variabel (Scimagojr, 2018).

**Tabel 4.1** Definisi Operasional Variabel

	Variabel	Definisi Operasional Variabel
Variabel Dependen	SJR Best Quartile	Memiliki empat kelompok (Q1, Q2, Q3, Q4) berdasarkan peringkat SJR yang didapat.
Variabel Independen	SJR	Jumlah rata-rata kutipan terimbang yang diterima pada tahun yang dipilih.
	H index	Jumlah artikel jurnal (h) yang paling banyak dikutip ilmuwan dan jumlah kutipan yang telah mereka terima dalam publikasi lain.
	Total Doc (2016)	Jumlah dokumen yang layak dan tidak layak yang dipublikasikan pada tahun 2016.
	Total Doc (3 year)	Jumlah dokumen layak dan tidak layak yang dipublikasikan dalam 3 tahun sebelumnya.
	Total Refs	Semua referensi bibliografi dalam jurnal pada periode yang dipilih
	Total Cites (3years)	Jumlah kutipan jurnal yang diterima oleh dokumen yang diterbitkan selama tiga tahun sebelumnya.
	Citable Docs (3years)	Jumlah dokumen layak yang diterbitkan oleh jurnal dalam tiga tahun sebelumnya.
	Cites Doc (2year)	Kutipan rata-rata dokumen dalam jangka waktu dua tahun dengan mempertimbangkan jumlah kutipan yang diterima oleh jurnal dalam tahun

	Variabel	Definisi Operasional Variabel
		berjalan ke dokumen yang diterbitkan dalam dua tahun sebelumnya.
	Ref Doc	Jumlah rata-rata referensi per dokumen pada tahun yang dipilih.
	Country	Nama negara asal peneliti yang jurnalnya terdapat di <i>scimago</i> .
	Categories	Kategori-kategori dari jurnal yang diterbitkan.

#### 4.4 Metode Pengumpulan Data

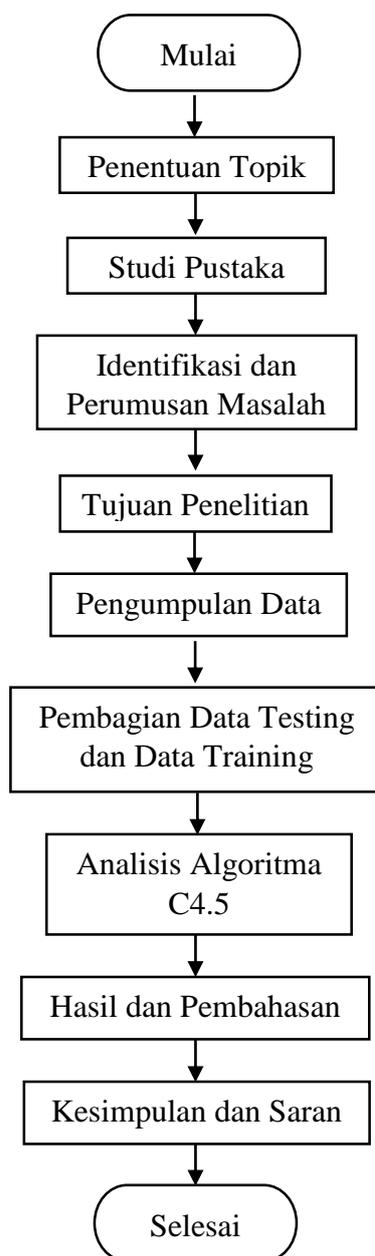
Pada penelitian ini, data yang digunakan adalah data sekunder yang didapatkan dari sumber-sumber yang telah ada. Data penelitian ini diperoleh dari *website scimagojr.com* pada tahun 2016 yang diakses pada tanggal 26 Maret 2018 (Scimagojr, 2018).

#### 4.5 Metode Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah analisis klasifikasi *Decision Tree* dengan menggunakan metode Algoritma C4.5. Klasifikasi merupakan suatu proses untuk menentukan sebuah model berdasarkan *data training* dan nilai kelas dari atribut target serta menggunakannya untuk mengklasifikasikan suatu data (Han, Kamber & Pei, 2011). Dengan analisis klasifikasi pada data jurnal ranking di *scimago*, maka para peneliti dapat dengan mudah mengetahui kategori jurnal apa saja yang memiliki kualitas tertinggi. Pada penelitian ini, alat yang digunakan untuk analisis klasifikasi (*classification*) adalah program R.

#### 4.6 Tahapan Penelitian *Decision Tree* Metode Algoritma C4.5

Tahapan atau langkah dalam penelitian ini digambarkan dalam *flowchart* pada Gambar 4.1 berikut ini :



**Gambar 4.1** *Flowchart* Penelitian

Berdasarkan *flowchart* penelitian diatas, dapat dilihat bahwa penelitian ini dimulai dengan menentukan topik penelitian terlebih dahulu. Kemudian dilakukan kajian studi pustaka sebagai landasan dalam penelitian. Selanjutnya dilakukan identifikasi dan perumusan masalah yang hendak diteliti, sehingga didapatkan tujuan penelitian yang akan dicapai. Langkah selanjutnya pengumpulan data yang mana data tersebut didapatkan dari *website* SJR. Setelah data terkumpul, maka selanjutnya melakukan pembagian data menjadi dua bagian yaitu *data training* dan *data testing*. Kemudian dilakukan analisis menggunakan metode algoritma C4.5.

Setelah didapatkan hasil dari analisis yang telah dilakukan, maka langkah selanjutnya dapat ditarik kesimpulan dan saran berdasarkan hasil tersebut.

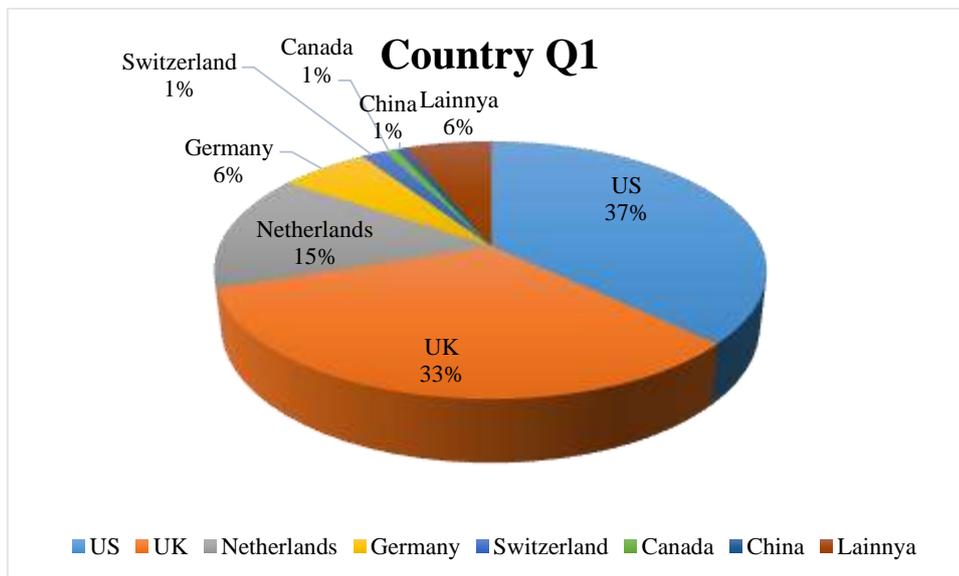
## BAB V

### HASIL DAN PEMBAHASAN

Pada bab ini, akan dilakukan analisis terhadap studi kasus *ranking* jurnal ilmiah menurut SJR (*Scimago Journal & Country Rank*) dari teori yang telah dibahas pada bab sebelumnya. Analisis yang digunakan dalam penelitian ini adalah klasifikasi menggunakan metode *Decision Tree*. Proses analisis dalam penelitian ini menggunakan bantuan *software Microsoft Excel 2013, R 3.4.4* dan *Tableau Public 10.3*.

#### 5.1 Analisis Deskriptif

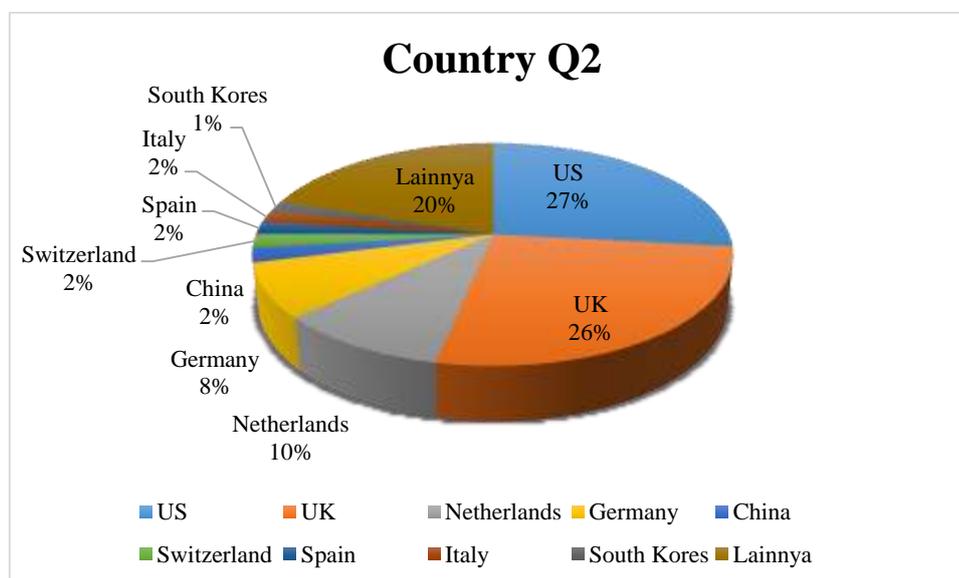
Analisis deskriptif dalam penelitian ini digunakan untuk melihat gambaran secara umum informasi tentang *Scimago Journal & Country Rank* berdasarkan data *journal country ranking* dari *website Scimagojr*, yang memiliki beberapa aspek diantaranya negara-negara yang masuk ke dalam kategori Q1, Q2, Q3, dan Q4 pada tahun 2016.



**Gambar 5.1** Presentase Negara Pada Q1

Berdasarkan Gambar 5.1 dapat diketahui bahwa negara yang memiliki presentase tertinggi pada kategori Q1 adalah *United States* dengan nilai 37%. Dari

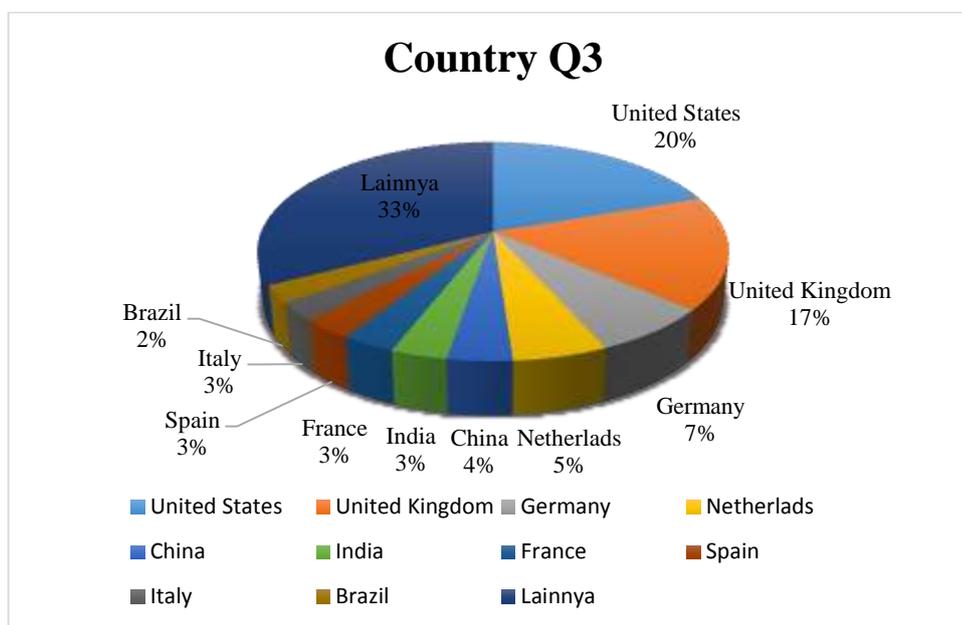
6.757 data jurnal yang ada, *United States* memiliki 2.491 jurnal pada kategori Q1 dan dari 105 negara yang terdapat di SJR, hanya 55 negara saja yang berhasil masuk ke dalam kategori Q1 ini. Dalam hal ini berdasarkan penilaian dari SJR dapat menggambarkan bahwa jurnal yang berasal dari *United States* memiliki kualitas yang tertinggi diantara jurnal dari negara lainnya, karena jurnal-jurnal yang masuk ke dalam kategori Q1 merupakan jurnal yang memiliki sisi kualitas tertinggi. Pada kelas Q1 ini, untuk negara Indonesia sendiri belum termasuk kedalam kuartil yang berarti bahwa jurnal-jurnal yang berasal dari Indonesia belum memiliki kualitas yang terbaik dibanding negara-negara maju yang lain.



**Gambar 5.2** Presentase Negara Pada Q2

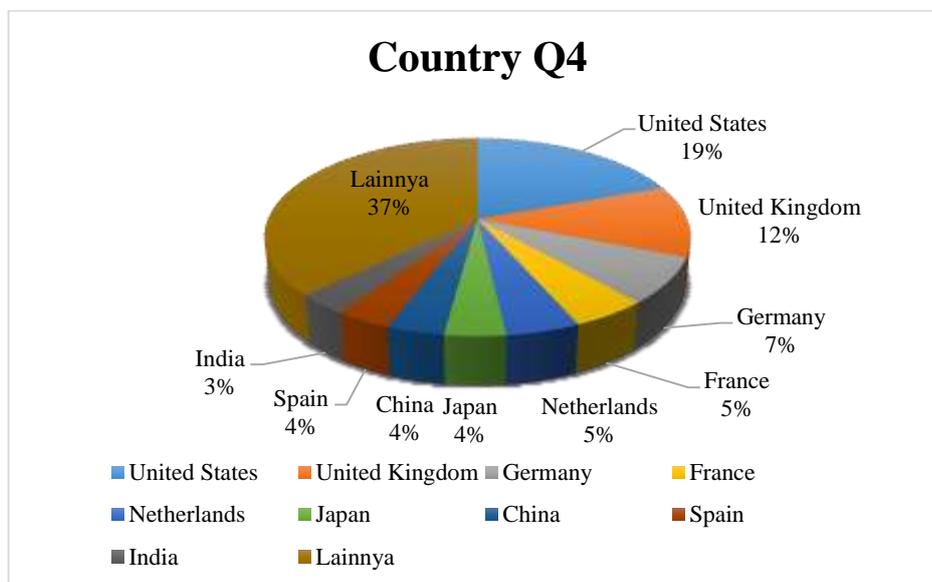
Dari Gambar 5.2 dapat diketahui bahwa negara yang memiliki presentase tertinggi pada kategori Q2 juga berasal dari *United States* dengan nilai 27%. Dari 6.032 data jurnal yang ada, *United States* memiliki 1.613 jurnal pada kategori Q2 dan dari 105 negara yang terdapat di SJR, hanya 71 negara saja yang berhasil masuk ke dalam kategori Q2 ini. Dalam hal ini menggambarkan bahwa jurnal yang berasal dari *United States* juga memiliki kualitas yang tertinggi diantara jurnal dari negara lainnya, karena pada kategori Q2 ini jurnal-jurnal yang masuk merupakan jurnal yang memiliki sisi kualitas tertinggi kedua setelah Q1. Pada kelas Q2, terdapat 2 jurnal ilmiah yang berasal dari negara Indonesia, yang berarti bahwa 2 jurnal

tersebut sudah memiliki kualitas yang cukup baik, walaupun kuantitasnya masih rendah dibanding dengan negara-negara maju yang lain.



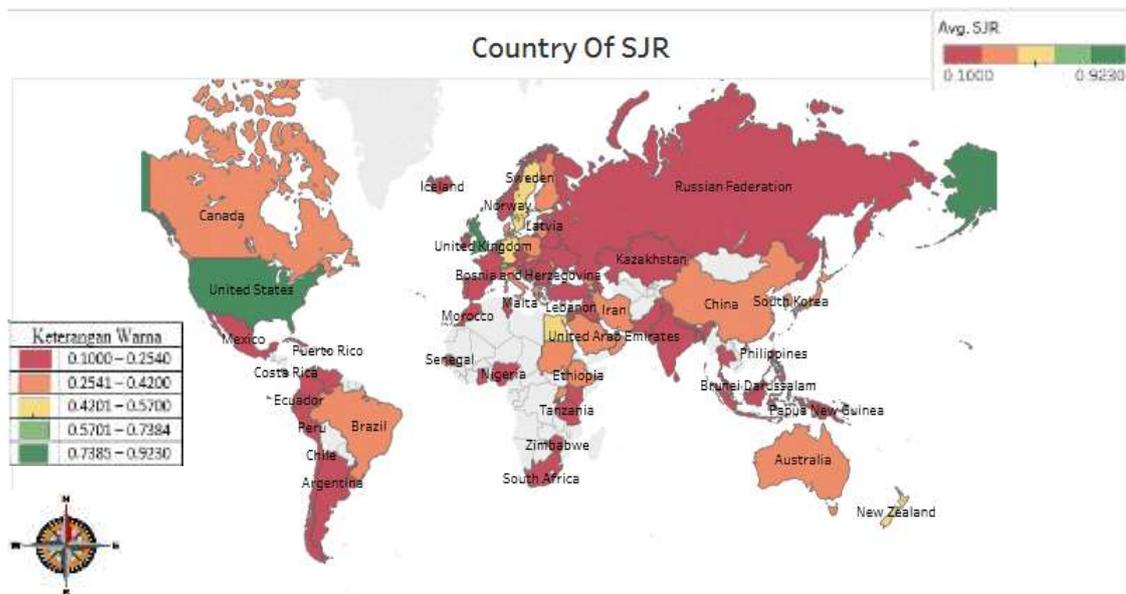
**Gambar 5.3** Presentase Negara Pada Q3

Dari Gambar 5.3 dapat diketahui bahwa negara yang memiliki presentase tertinggi pada kategori Q3 juga berasal dari *United States* dengan nilai 19,59%. Dari 5.721 data jurnal yang ada, *United States* memiliki 1.121 jurnal pada kategori Q3 dan dari 105 negara yang terdapat di SJR, hanya 89 negara saja yang berhasil masuk ke dalam kategori Q3 ini. Untuk negara Indonesia sendiri, jurnal ilmiah yang masuk pada kategori Q3 ini sebanyak 9 jurnal ilmiah. Jika dilihat dari banyaknya jurnal, maka dapat dilihat secara kasat mata bahwa jurnal dari negara Indonesia masih sangat jauh dibandingkan negara-negara maju pada umumnya, namun dari 23 total jurnal Indonesia dan 9 diantaranya sudah masuk di kategori Q3 yang berarti kualitas yang dihasilkan dari jurnal Indonesia sudah cukup baik. Pada kategori Q3 ini jurnal-jurnal yang masuk merupakan jurnal yang memiliki sisi kualitas tertinggi ketiga setelah Q1 dan Q2.



**Gambar 5.4** Presentase Negara Pada Q4

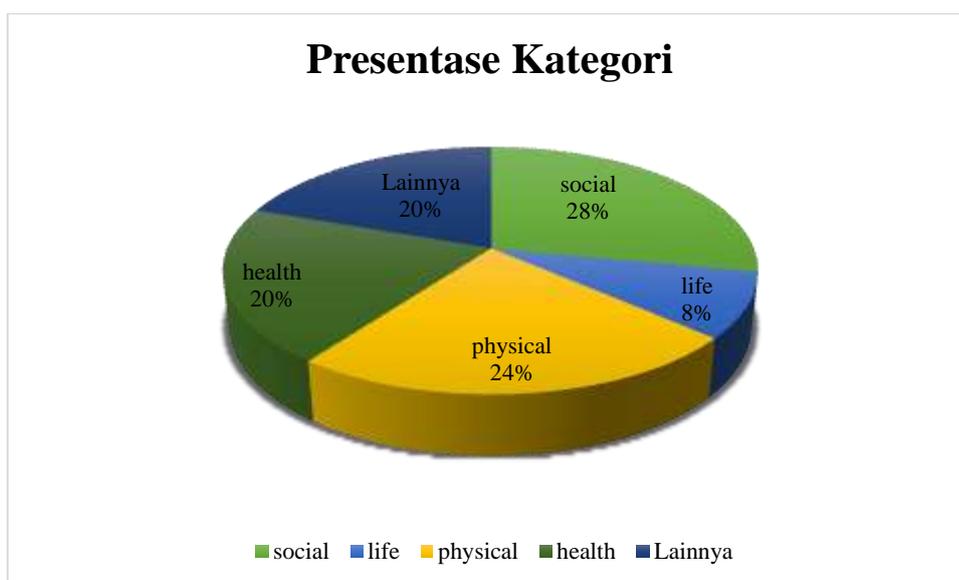
Dari Gambar 5.4 dapat diketahui bahwa negara yang memiliki presentase tertinggi pada kategori Q4 juga berasal dari *United States* dengan nilai 19,32%. Dari 5.419 data jurnal yang ada, *United States* memiliki 1.047 jurnal pada kategori Q4 dan dari 105 negara yang terdapat di SJR, hanya 93 negara saja yang berhasil masuk ke dalam kategori Q4 ini, salah satunya yaitu Indonesia yang mana jumlah jurnal Indonesia yang masuk pada kategori ini sebanyak 12 jurnal dari 23 total jurnal yang dimiliki. Pada kategori Q4 ini jurnal-jurnal yang masuk merupakan jurnal yang memiliki sisi kualitas terendah dari kuartil yang lainnya.



**Gambar 5.5** Persebaran Negara Berdasarkan Nilai SJR

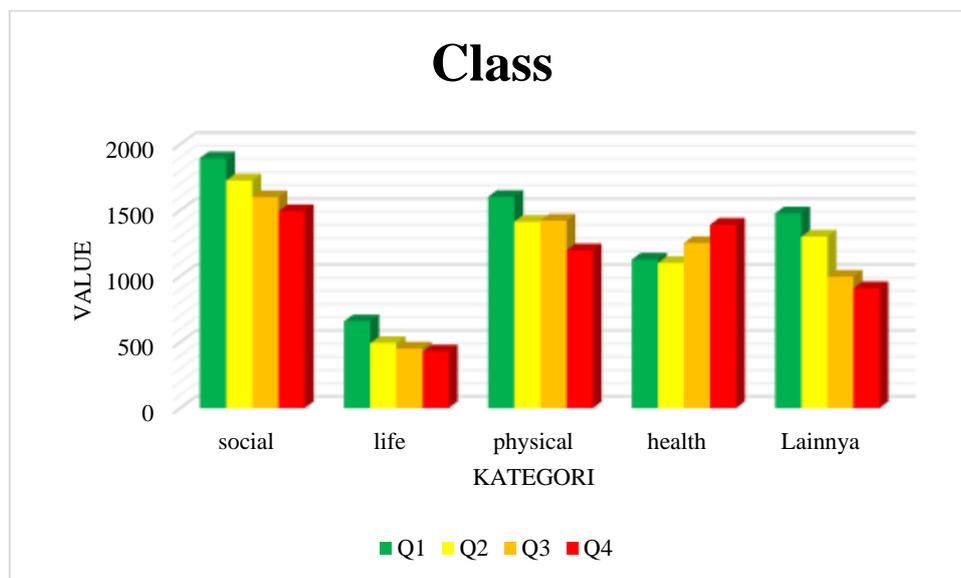
Gambar 5.5 diatas menunjukkan persebaran negara-negara yang masuk ke dalam *Scimago Journal & Country Ranking* berdasarkan nilai rata-rata SJR yang didapatkan dari jurnal-jurnal yang telah terindeks oleh *Scopus*. Pada gambar tersebut dibuat lima level gradiasi warna untuk membedakan besar-kecilnya nilai rata-rata SJR yang didapat suatu negara. Pada gradiasi warna pertama yaitu warna merah tua, gradiasi warna kedua yaitu warna oranye, gradiasi warna ketiga yaitu warna kuning, gradiasi warna keempat yaitu warna hijau, dan gradiasi warna yang kelima yaitu warna hijau tua. Berdasarkan warna gradiasi tersebut, maka negara yang memiliki gradiasi warna merah tua adalah kategori negara yang memiliki nilai rata-rata SJR nya paling rendah dan yang memiliki gradiasi warna hijau tua berarti negara tersebut berada pada kategori yang memiliki nilai rata-rata SJR yang tertinggi. Gradiasi warna yang pertama memiliki nilai rata-rata SJR dari 0,1000 – 0,2540. Gradiasi warna kedua memiliki rata-rata nilai SJR dari 0,2541 – 0,4200. Gradiasi warna ketiga memiliki nilai rata-rata SJR dari 0,4201 – 0,5700. Gradiasi warna keempat memiliki nilai rata-rata SJR dari 0,5701 – 0,7384. Gradiasi warna yang kelima memiliki nilai rata-rata SJR dari 0,7385 – 0,9230. Jika dilihat pada gambar 5.5 diatas, maka dapat diketahui bahwa negara *United States* dan *United Kingdom* mempunyai gradiasi warna hijau tua yang berarti bahwa negara tersebut masuk ke dalam kategori negara yang memiliki nilai rata-rata SJR yang tertinggi,

sedangkan untuk negara yang masuk ke dalam gradiasi warna merah tua salah satunya adalah negara Indonesia. Hal ini disebabkan karena masih sedikitnya jurnal ilmiah yang dihasilkan Indonesia dibandingkan dengan negara-negara maju lainnya. Rata-rata nilai SJR yang didapatkan Indonesia hanya sebesar 0,1615. Jika dibandingkan dengan negara tetangga seperti Thailand maka Indonesia masih dibawah nilai rata-rata SJR dari Thailand.



**Gambar 5.6** Presentase Kategori Jurnal

Gambar 5.6 diatas menunjukkan bahwa kategori jurnal yang paling banyak di SJR adalah kategori *Social Science* yaitu sebesar 28%, kemudian kategori *Physical Science* sebesar 24%, kategori *Health Science* sebesar 20%, kategori *Lifie Science* sebesar 8%, dan sisanya sebanyak 20% berasal dari gabungan beberapa kategori tersebut. Dari total keseluruhan data jurnal yang terdapat di SJR sebanyak 23.929 jurnal, kategori *Social Science* memiliki 6.712 jurnal, kategori *Life Science* 2.033 jurnal, kategori *Physical Science* 5.632 jurnal, kategori *Health Science* 4.870 jurnal, dan sisanya berasal dari gabungan beberapa kategori yang ada.



**Gambar 5.7** Perbandingan Kategori Berdasarkan *Quartile*

Dilihat dari Gambar 5.7 yang menunjukkan perbandingan kategori jurnal berdasarkan *Quartile*. Jenis kategori yang paling banyak disetiap *Quartile* yaitu kategori *Social Science*, sebanyak 1.894 jurnal pada Q1, 1.725 jurnal pada Q2, 1.600 jurnal pada Q3, dan 1.493 jurnal pada Q4. Untuk kategori *Life Science* paling banyak terdapat pada Q1 yaitu 659 jurnal, 493 jurnal pada Q2, 452 pada Q3, dan 429 pada Q4. Kategori *Physical Science* paling banyak juga pada Q1 yaitu 1.601 jurnal, 1.413 jurnal pada Q2, 1.421 jurnal pada Q3, 1.197 jurnal pada Q4. Pada kategori *Health Science* paling banyak terdapat pada Q4 yaitu 1.391 jurnal, kemudian disusul Q3 sebanyak 1.252 jurnal, selanjutnya Q1 sebanyak 1.126 jurnal, dan terakhir Q2 sebanyak 1.101 jurnal.

## 5.2 Analisis Regresi

**Tabel 5.1** Uji Overall

ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10861,617	9	1206,846	1497,445	,000 <sup>b</sup>
	Residual	19277,204	23919	,806		
	Total	30138,822	23928			

a. Dependent Variable: Quartile

b. Predictors: (Constant), Ref\_Doc, Citable, SJR, H\_index, Refs, Cites, Cites\_Doc, Doc\_2016, Doc\_3years

Berdasarkan table diatas, akan dibuktikan apakah pada data diatas variabel independen (Ref\_Doc, Citable, SJR, H\_index, Refs, Cites, Cites\_Doc, Doc\_2016, Doc\_3years) memiliki pengaruh yang signifikan terhadap variabel dependen (Quartile).

a. Hipotesis

$H_0$  : Tidak ada pengaruh secara signifikan antara semua variabel independen terhadap variabel dependen.

$H_1$  : Ada pengaruh secara signifikan antara semua variabel independen terhadap variabel dependen

b. Tingkat Signifikansi

$$\alpha = 5\% = 0,05$$

c. Daerah Kritis

Tolak  $H_0$  jika  $\text{Sig.} < \alpha$

d. Statistik Uji

$$\text{Sig.} = 0,000$$

$$\alpha = 0,05$$

Sehingga  $\text{Sig.} < \alpha (0,05)$

e. Keputusan

Karena nilai  $\text{Sig.} < \alpha$  maka keputusannya adalah tolak  $H_0$ .

f. Kesimpulan

Dengan tingkat kepercayaan 95%, variabel independen diatas diketahui secara keseluruhan memiliki hubungan yang signifikan dengan variabel dependen.

Dari hasil uji hipotesis diatas didapatkan bahwa secara keseluruhan variabel independen memiliki pengaruh secara signifikan terhadap variabel dependen, oleh karena itu maka diperlukannya uji hipotesis secara parsial untuk mengetahui lebih jelas pengaruh dari masing-masing variabel independen terhadap variabel dependen tersebut.

**Tabel 5.2 Uji Parsial****Coefficients<sup>a</sup>**

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	3,123	,010		321,788	,000
SJR	,009	,009	,010	,990	,322
H_index	-,012	,000	-,453	-55,490	,000
Doc_2016	-8,912E-5	,000	-,026	-1,451	,147
Doc_3years	,000	,000	,436	7,331	,000
Refs	-1,159E-5	,000	-,136	-7,471	,000
Cites	8,418E-5	,000	,258	24,865	,000
Citable	,000	,000	-,435	-7,660	,000
Cites_Doc	-,090	,005	-,174	-16,501	,000
Ref_Doc	-,008	,000	-,195	-33,738	,000

a. Dependent Variable: Quartile

Berdasarkan table diatas, akan dibuktikan apakah pada data diatas variabel independen (Ref\_Doc, Citable, SJR, H\_index, Refs, Cites, Cites\_Doc, Doc\_2016, Doc\_3years) memiliki pengaruh yang signifikan terhadap variabel dependen (Quartile).

- Hipotesis

$H_0$  : Secara parsial tidak ada pengaruh signifikan antara masing-masing variabel independen terhadap variabel dependen.

$H_1$  : Secara parsial ada pengaruh signifikan antara masing-masing variabel independen terhadap variabel dependen

Tingkat Signifikansi	Daerah Kritis	Statistik Uji	Keputusan
0,05	Sig. < $\alpha$	0,322 > 0,05	Gagal Tolak $H_0$
0,05	Sig. < $\alpha$	0,000 < 0,05	Tolak $H_0$
0,05	Sig. < $\alpha$	0,147 > 0,05	Gagal Tolak $H_0$
0,05	Sig. < $\alpha$	0,000 < 0,05	Tolak $H_0$
0,05	Sig. < $\alpha$	0,000 < 0,05	Tolak $H_0$

0,05	Sig. < $\alpha$	0,000 < 0,05	Tolak $H_0$
0,05	Sig. < $\alpha$	0,000 < 0,05	Tolak $H_0$
0,05	Sig. < $\alpha$	0,000 < 0,05	Tolak $H_0$
0,05	Sig. < $\alpha$	0,000 < 0,05	Tolak $H_0$

- Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan dengan menggunakan tingkat kepercayaan 95%, sehingga dapat ditarik kesimpulan bahwa variabel SJR dan variabel Doc\_2016 tidak memiliki pengaruh yang signifikan terhadap variabel dependen.

Penggunaan analisis regresi pada penelitian hanya untuk mengetahui variabel independen mana saja yang memiliki pengaruh atau tidak terhadap variabel dependen. Setelah dilakukan analisis regresi secara parsial, maka diperoleh 2 variabel independen yang tidak memiliki pengaruh secara signifikan terhadap variabel dependen. Walaupun variabel independen tersebut tidak berpengaruh terhadap variabel dependen, namun variabel tersebut masih digunakan dalam penelitian ini karena secara *overall* semua variabel independen tersebut memiliki pengaruh yang signifikan terhadap variabel dependen.

### 5.3 Klasifikasi *Decision Tree* Dengan Algoritma C4.5

#### 5.1 Penentuan Atribut Pada *Node*

Algoritma C4.5 adalah salah satu metode pembentukan pohon keputusan, sehingga adanya penentuan atribut pada *node* pohon keputusan tersebut nantiya. Atribut pada *node* ditentukan dengan menggunakan nilai gain rasio tertinggi yang akan menjadi *node*. Pohon keputusan dimulai dengan memilih atribut sebagai akar (*root node*), *root node* adalah *level node* teratas di sebuah pohon keputusan (*decision tree*) yang merupakan atribut dengan pengaruh terbesar. Setelah menentukan *root node*, selanjutnya yaitu menentukan *node* cabang dengan menggunakan gain rasio. Berikut ini merupakan contoh perhitungan untuk mendapatkan nilai gain rasio masing-masing atribut, sehingga dapat ditentukan untuk *node* pohon keputusan :

a. Mencari Nilai Entropy

Nilai Entropy yang dicari adalah nilai entropy total dan nilai entropy masing-masing kategori untuk semua variabel. Berikut merupakan contoh perhitungan untuk mencari nilai entropy total dan entropy variabel, variabel yang akan dijadikan contoh adalah variabel *Quartile* :

$$Entropy = - \sum_{i=1}^n p_i \log_2(p_i)$$

$$\begin{aligned} Entropy \text{ (Total)} &= - \left( \left( \frac{6757}{23929} \log_2 \left( \frac{6757}{23929} \right) \right) + \left( \frac{6032}{23929} \log_2 \left( \frac{6032}{23929} \right) \right) \right. \\ &\quad \left. + \left( \frac{5721}{23929} \log_2 \left( \frac{5721}{23929} \right) \right) + \left( \frac{5419}{23929} \log_2 \left( \frac{5419}{23929} \right) \right) \right) \\ &= 1.95 \end{aligned}$$

b. Mencari Nilai Gain

Setelah mendapatkan nilai entropy dari semua variabel, kemudian mencari nilai gain untuk setiap variabel.

c. Mencari Nilai Gain Rasio

## 5.2 Pohon Keputusan Algoritma C4.5

Berdasarkan lampiran 3, maka didapatkan beberapa aturan seperti di Tabel 5.3 dibawah. Dari pohon klasifikasi yang dihasilkan dapat dilihat bahwa variabel yang menjadi *node* paling awal (*root node*) adalah variabel H\_index, karena variabel H\_index memiliki nilai *Gain Ratio* yang paling tinggi diantara variable lainnya dan merupakan variabel yang sangat berpengaruh. Selain variabel H\_index, variabel lain yang memiliki pengaruh yaitu variabel SJR, Ref\_Doc.

**Tabel 5.3** Segmentasi Pohon Keputusan Algoritma C4.5

Segmen	Keterangan
1	H_index dengan nilai $\leq 39$ , SJR $\leq 0.16$ , SJR $\leq 0.129$ , SJR $\leq 0.108$ , dan SJR $\leq 0.1$
:	:
57	H_index dengan nilai $> 39$ , SJR $\leq 0.704$ , dan SJR $\leq 0.191$ .
:	:
66	H_index dengan nilai $> 39$ , SJR $> 0.704$ , SJR $> 1.039$ , SJR $> 1.422$

Jika dilihat dari Tabel 5.3 diatas, maka dapat diketahui bahwa analisis pohon keputusan dengan menggunakan data *journal ranking scimago* sebanyak 23.292 data dapat menghasilkan pohon keputusan yang memiliki 66 terminal *node*. Terlihat dari segmen pertama yang menjadi *root* adalah variabel H\_index, dimana ketentuan untuk menjadikan suatu variabel menjadi *root* dari pohon keputusan itu sendiri adalah variabel yang memiliki nilai *gain* terbesar dari variabel-variabel yang lainnya. Aturan yang akan dipakai untuk melihat prediksi kelas yang dihasilkan menggunakan klasifikasi algoritma C4.5 ini adalah aturan segmentasi yang pertama dan segmentasi ke-57.

Berikut merupakan beberapa prediksi kelas yang didapatkan berdasarkan nilai yang terbentuk pada aturan-aturan diatas.

```

> SJR = c(0.1)                                > SJR = c(0.191)
> H_index = c(39)                             > H_index = c(40)
> x_test = data.frame(SJR, H_index)           > x_test = data.frame(SJR, H_index)
> predicted = predict(fit, x_test)            > predicted = predict(fit, x_test)
> predicted                                     > predicted
      Q1      Q2      Q3      Q4                Q1      Q2      Q3      Q4
1 0.01350566 0.08350276 0.213041 0.6899506  1 0.06899152 0.19623 0.6137606 0.1210179

```

**Gambar 5.8** Prediksi Kelas

Berdasarkan gambar 5.8, dengan menggunakan aturan segmen ke-1 maka didapatkan hasil prediksi untuk data yang memiliki nilai SJR 0.1, H index 39 maka data tersebut akan diprediksi masuk kedalam kelas Q4, karena data tersebut memiliki nilai probabilitas yang paling tinggi di kelas Q4. Untuk contoh yang kedua yaitu menggunakan aturan segmen yang ke-57 yang memiliki nilai SJR 0.191, H index 40, maka data tersebut akan diprediksi masuk kedalam kelas Q3, karena nilai

probabilitas dari prediksi data tersebut paling besar terdapat pada kelas Q3. Untuk melihat prediksi kelas yang lain, maka peneliti dapat memasukkan nilai berapa saja yang akan digunakan dengan menggunakan rumus seperti diatas dan selanjutnya akan didapatkan hasil prediksi dari kelas tersebut.

Pada penelitian ini menggunakan evaluasi model dengan *confusion matrix* untuk mengetahui hasil akurasi dari klasifikasi, sedangkan untuk melakukan evaluasi model pada penelitian ini dilakukan dengan dengan membuat 10 buah *machine learning* untuk menentukan nilai akurasi prediksi terbaik. Adapun hasil dari masing-masing percobaan *machine learning* menggunakan algoritma C4.5 adalah sebagai berikut :

**Tabel 5.4** Perbandingan Nilai Akurasi *Machine Learning*

<b>Machine Learning</b>	<b>Akurasi Metode</b>
<b>Machine Learning 1</b>	69,91%
<b>Machine Learning 2</b>	70,35%
<b>Machine Learning 3</b>	69,89%
<b>Machine Learning 4</b>	70,97%
<b>Machine Learning 5</b>	69,93%
<b>Machine Learning 6</b>	70,17%
<b>Machine Learning 7</b>	70,32%
<b>Machine Learning 8</b>	69,42%
<b>Machine Learning 9</b>	69,44%
<b>Machine Learning 10</b>	69,86%

Berdasarkan Tabel 5.4 diatas, dapat dilihat dari 10 percobaan yang telah dilakukan menggunakan algoritma C4.5 pada data SJR, percobaan pada *machine learning* ke-4 menghasilkan nilai akurasi yang tertinggi yakni sebesar 70,97%. Hasil perhitungan tingkat akurasi tersebut diperoleh berdasarkan jumlah data uji yang terklasifikasi benar dibandingkan dengan total semua data uji.

Untuk menguji performa dari *machine learning* dalam melakukan klasifikasi, maka dilakukan *cross validation* menggunakan *10-fold cross validation* dengan hasil rata-rata akurasi yang diperoleh sebesar 70,03%. Untuk memudahkan proses perhitungan akurasi, maka digunakan *confusion matrix* untuk mengetahui jumlah data uji yang terklasifikasi dengan benar dan jumlah dat uji yang salah

pengklasifikasiannya. Adapun *confusion matrix* yang diperoleh dari *machine learning* ke-4 dapat dilihat pada tabel 5.5 berikut.

**Tabel 5.5** *Confusion Matrix*

	Q1	Q2	Q3	Q4	Total
Q1	1.636	351	20	0	2.007
Q2	241	1.093	328	20	1.682
Q3	92	250	996	237	1.575
Q4	17	144	389	1.381	1.931
Total	1.986	1.838	1.733	1.638	7.195

Tabel 5.5 digunakan untuk melihat berapa akurasi dari analisis menggunakan algoritma C4.5 dengan *data testing*. Perhitungan dalam mencari nilai akurasi dengan menggunakan tabel *confusion matrix* tersebut yaitu :

$$\text{Akurasi} = \frac{\sum \text{Prediksi Benar}}{\sum (\text{Benar} + \text{Salah})} = \frac{5.106}{7.195} = 0,7097 = 70,97\%$$

$$\text{Error Rate} = \frac{\sum \text{Prediksi Salah}}{\sum (\text{Benar} + \text{Salah})} = \frac{2.089}{7.195} = 0,2903 = 29,03\%$$

Berdasarkan perhitungan, didapatkan bahwa nilai keakuratan algoritma C4.5 dalam menganalisis data *Scimago Journal & Country Ranking* adalah sebesar 70,97%, yang artinya dari total data uji pada Q1 sebanyak 2.007 maka yang terklasifikasi dengan benar sebanyak 1.636 data, pada Q2 dari 1.838 total data Q2 maka yang terklasifikasi dengan benar sebanyak 1.093, untuk Q3 dengan total data sebanyak 1.733 yang terklasifikasi dengan benar sebanyak 996 data, sedangkan untuk total data dari Q4 sebanyak 1.638 yang terklasifikasi sebanyak 1.381 data, dan tingkat kesalahan dalam algoritma C4.5 untuk data tersebut adalah sebesar 29,03%.

## **BAB VI**

### **PENUTUP**

#### **6.1 Kesimpulan**

Dari hasil pembahasan yang telah dijabarkan diatas, maka dapat ditarik kesimpulan bahwa :

1. Gambaran umum yang didapat berdasarkan data SJR diketahui bahwa jurnal yang banyak terindeks di SJR adalah jurnal yang berasal dari negara *United States*. Untuk negara Indonesia sendiri kuantitas jurnal yang dihasilkan masih sangat sedikit dan kualitasnya pun belum ada yang masuk kedalam kategori kualitas tertinggi masih berada di kategori Q2. Jika dilihat dari kategori subjek, maka yang paling banyak mendominasi adalah kategori subjek *Social Science*, kemudian diikuti *Physical Science*, *Health Science*, dan terakhir *Life Science*.
2. Dari analisis klasifikasi menggunakan *Decision Tree* Algoritma C4.5 didapatkan pohon klasifikasi yang memiliki 66 *node* dengan 10 variabel yang diujikan.
3. Berdasarkan model klasifikasi yang telah digunakan, maka didapatkan nilai akurasi untuk *data testing* sebesar 70,97 %.

#### **6.2 Saran**

Berdasarkan hasil analisis dan kesimpulan, penulis memberikan beberapa beberapa saran sebagai berikut :

1. Untuk penelitian selanjutnya, diharapkan dapat menggunakan metode *decision tree* lainnya seperti metode *Random Forest* atau metode yang lain untuk membandingkan tingkat akurasi klasifikasi agar mendapatkan model yang terbaik.
2. Bisa menggunakan dat publikasi ilmiah internasional yang lain seperti ISI Thomson, untuk mengetahui perbandingan jurnal yang seperti apa yang terindeks di masing-masing publikasi.

## DAFTAR PUSTAKA

- Adnan, Z., dan Zifirdaus, I. (2005). *Merebut Hati Audiens Internasional: Strategi Ampuh Meraih Publikasi di Jurnal Ilmiah*. Jakarta: Gramedia.
- Astuti, F. H. (2013). *Data Mining*. Yogyakarta: Andi Offset.
- Berry, M. J., dan Linoff, G. S. (2004). *Data Mining Techniques For Marketing Sales, Customer, Relationship Management*. Wiley Publishing, Inc.
- Elsevier.(2018).*About Scopus*. <http://www.elsevier.com/solutions/scopus>. Akses, 02 April 2018
- Elsevier.(2018).*About Elsevier*. <http://www.elsevier.com/publisher-relations>. Akses 02 April 2018
- Estoatnowo, D. (2017). *Klasifikasi Status Kelulusan Mahasiswa Menggunakan Metode CHAID Dan Algoritma C4.5*. Yogyakarta: Universitas Islam Indonesia.
- Fauzie, R.(2010).*Pengenalan Citra Wajah Menggunakan Algoritma VF15 dengan Praproses Principal Component Analysis*.Bogor, Institut Teknologi Pangan.
- Gorunescu, F. (2011). *Data Mining : Concept, Model and Techniques*. Berlin: Springer.
- Han, J., Kamber, M., dan Pei, J. (2001). *Data Mining Concept and Techniques Chapter 1*. San Fransisco: Morgan Kaufman Publisher.
- Harryanto, Fendy F., dan Hansun, S.(2017).Penerapan Algoritma C4.5 Untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE.*Jurnal Jatisi, III (2)*.
- Haryati, S., Sudarsono, A., dan Suryana, E. (2015). Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5

- (Studi Kasus : Universitas Dehasen Bengkulu). *Jurnal Media Informa*, 11 (2), 130-138.
- Ian, H., Frank, E., dan Mark, A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques - 3rd ed.* Burlington: Elsevier.
- Indranandita, A., Susanto, B., dan Rachmat, A. (2008). Sistem Klasifikasi dan Pencarian Jurnal Dengan Menggunakan Metode Naive Bayes dan Vector Space Model. *Jurnal Informatika*, IV.
- James, G., Witten, D., Hastie, T., dan Tibshirani, R.(2014).*An Introduction to Statistical Learning With Applications in R.*New York: Springer.
- Jayanegara, A. (2015). *Kiat Diterima Publikasi di Jurnal Internasional.* Bogor: Institut Pertanian Bogor.
- Julianto, W., Yunitarini, R., dan Sophan, M. K. (2014). Implementasi Data Mining Dengan Algoritma C4.5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal ULTIMATICS*, VI (1).
- Karim, M., dan Rahman, R. M.(2013).Decision Tree and Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing.*Journal of Software Engineering and Application.*196-206.
- Kohavi, R., dan Provost, F. (1998). Guest Editor's Introduction. *On Applied Research in Machie Learning*, 127-132.
- Kriyantono, R. (2015). Strategi Praktis Menulis untuk Jurnal Internasional Bereputasi.
- Kurniawan, M. F., dan Iwandari. (2017, April). Komparasi Algoritma Data Mining Untuk Klasifikasi Penyakit Kanker Payudara. *IC-Tech*, XII.
- Larose, D. T., dan Larose, C. D. (2005). *Discovering Knowledge In Data : An Introduction to Data Mining.* Canada: John Willey & Sons.Inc.

- Mashlahah, S. (2013). *Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5*. Malang: UIN Maulana Malik Ibrahim.
- Oktaviana, A. R. (2016). *Penerapan Data Mining Klasifikasi Pola Nasabah Menggunakan Algoritma C4.5 Pada Bank BRI Batang*. Semarang: Universitas Dian Nuswantoro.
- Paratu, L.G. (2013). *Analisis Pada Pelanggan Telekomunikasi Menggunakan Algoritma C4.5*. Skripsi. Program Studi Statistika Jurusan Matematika FMIPA UGM. Yogyakarta.
- Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi Offset.
- Pratiwi. (2010). *Pengembangan Model Pengenalan Wajah Dengan Jarak Euclid Pada Ruang Eigen Dengan 2DPCA*. Bogor, Program Pascasarjana, Institut Pertanian Bogor.
- Pratomo, D. (2015). *Penulisan Jurnal Internasional Bereputasi : Pelatihan dan Konsinyering Penulisan Jurnal*. Malang: Universitas Brawijaya.
- Rani, L. N. (2015). *Klasifikasi Nasabah Menggunakan Algoritma C4.5 sebagai Dasar Pemberian Kredit*. *Jurnal KomTekInfo Fakultas Ilmu Komputer*, 2 (2).
- Rianto, B. (2016). *Implementasi dan Perbandingan Metode Prapemrosesan Pada Analisis Sentimen Gubernur DKI Jakarta Menggunakan Metode Support Vector Machine dan Naive Bayes*. Skripsi, Universitas Gadjah Mada Yogyakarta: Program Studi Ilmu Komputer FMIPA.
- Ristekdikti. (2013). *Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 17 Tahun 2013*. Jakarta. dari <http://www.kelembagaan.ristekdikti.go.id>. Akses, 10 April 2018

- Rasywir, E., dan Purwarianti, A. (2015). Eksperimen Pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin. *Jurnal Cybermatika*, 3 No. 2.
- Scimago Journal & Country Rank. (2018). Diambil kembali dari <http://www.scimagojr.com>. Akses, 26 Maret 2018)
- Sekaran, U. (2011). *Research Methods for Business Edisi I and 2*. Jakarta: Salemba Empat.
- Sharma, A., dan Srivastava, A.(2016).Understanding Decision Tree Algorithm by Using R Programming Language.*ACEIT Conference Proceeding*.
- Soewono, R., Gemowo, R., dan Sasongko, P. S. (2014). Sistem Pakar Identifikasi Modalitas Belajar Siswa Dengan Implementasi Algoritma C4.5. *Jurnal Sistem Informasi Bisnis*, 1, 20-27.
- Sunjana. (2010). Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree. *SNATI*.
- Supriyanti, W., Kusriani, dan Amborowati, A.(2016).Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa.*Jurnal INFORMA*, 1 (3).
- Suryoputro, G., Rudi, S., dan Sya'ban, A. (2012). *Menulis Artikel Untuk Jurnal Ilmiah*. Jakarta: Uhamka Press.
- Turban, E.(2005). *Decision Support System And Intelligent System*. Upper Saddle River, New Jersey USA: Prentice Hall.
- Wikipedia.(2017).*Analisis Regresi*.  
[https://id.m.wikipedia.org/wiki/Analisis\\_regresi](https://id.m.wikipedia.org/wiki/Analisis_regresi). Akses, 20 Mei 2018

# LAMPIRAN

## LAMPIRAN

**Lampiran 1. Dataset SJR**

Rank	Title	Type	Issn	SJR	SJR_Best_Quartile	H_index	Total_Docs_2016	Total_Docs_3years	Total_Refs	Total_Cites_3years	Citable_Docs_3 years	Cites_Doc_2years	Ref_Doc	Country	Categories
1	CA - A Cancer Journal for Clinicians	journal	ISSN 15424863, 00079235	39,285	Q1	131	43	141	3503	11929	118	128,75	81,47	United States	Hematology (Q1); Oncology (Q1)
2	Nature Reviews Genetics	journal	ISSN 14710056, 14710064	33,238	Q1	292	166	615	8029	7131	183	39,69	48,37	United Kingdom	Genetics (Q1); Genetics (clinical) (Q1); Molecular Biology (Q1)
3	Nature Reviews	journal	ISSN 147	29,692	Q1	316	146	581	7719	8256	195	36,47	52,87	United Kingdom	Immunology (Q1); Medicine

Rank	Title	Type	Issn	SJR	SJR_Best_Quartile	H_index	Total_Docs_2016	Total_Docs_3years	Total_Refs	Total_Cites_3years	Citable_Docs_3 years	Cites_Doc_2years	Ref_Doc	Country	Categories
	Immunology		41733												(miscellaneous) (Q1)
4	Nature Reviews Molecular Cell Biology	journal	ISSN 14710072, 14710080	29,656	Q1	352	152	535	9128	8150	214	45,11	60,05	United Kingdom	Cell Biology (Q1); Molecular Biology (Q1)
5	Annual Review of Immunology	journal	ISSN 07320582, 15453278	27,631	Q1	267	23	72	4155	2513	72	28,83	180,65	United States	Immunology (Q1); Immunology and Allergy (Q1)
6	Cell	journal	ISSN 00928674,	26,947	Q1	655	693	1885	29440	42666	1690	23,55	42,48	United States	Biochemistry, Genetics and Molecular Biology

Rank	Title	Type	Issn	SJR	SJR_Best_Quartile	H_index	Total_Docs_2016	Total_Docs_3years	Total_Refs	Total_Cites_3years	Citable_Docs_3 years	Cites_Doc_2years	Ref_Doc	Country	Categories
			10974172												(miscellaneous) (Q1)
7	Quarterly Journal of Economics	journal	ISSN 00335315314650	24,769	Q1	205	40	121	2401	986	120	7,45	60,03	United States	Economics and Econometrics (Q1)
8	Reviews of Modern Physics	journal	ISSN 0034686115390756	23,543	Q1	271	46	123	12278	4383	117	36,57	266,91	United States	Physics and Astronomy (miscellaneous) (Q1)
9	Annual Review of Astronomy	journal	ISSN 15454282006	22,491	Q1	151	18	43	4779	1512	42	30,53	265,5	United States	Astronomy and Astrophysics (Q1); Space and Planetary Science (Q1)

Rank	Title	Type	Issn	SJR	SJR_Best_Quartile	H_index	Total_Docs_2016	Total_Docs_3years	Total_Refs	Total_Cites_3years	Citable_Docs_3 years	Cites_Doc_2years	Ref_Doc	Country	Categories
	and Astrophysics		64146												
10	Nature Reviews Cancer	journal	ISSN 1474175X	21,53	Q1	355	115	513	9863	7806	207	32,32	85,77	United Kingdom	Cancer Research (Q1); Oncology (Q1)
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
23929	Economist	journal	ISSN 00130613	0,1	Q4	6	675	12204	0	2	9932	0	0	United Kingdom	Biochemistry (Q1) Business, Management and Accounting (miscellaneous) (Q4); Economics and Econometrics (Q4)

## Lampiran 2. Syntax Decision Tree Algoritma C4.5

```
data <- read.delim("clipboard")
scimago <- data
str(scimago)
set.seed(1234)
ind <- sample(2, nrow(scimago), replace=TRUE, prob=c(0.7, 0.3))
trainData <- scimago[ind==1,]
testData <- scimago[ind==2,]
install.packages("party")
library(party)
myFormula <- SJR_Best_Quartile ~ SJR + H_index + Total_Docs_2016
+ Total_Docs_3years + Total_Refs + Total_Cites_3years +
Citable_Docs_3years + Cites_Doc_2years + Ref_Doc
scimago_ctree <- ctree(myFormula, data=trainData)
tab <- table(predict(scimago_ctree),
trainData$SJR_Best_Quartile)
accuracy <- sum(tab[1,1]+tab[2,2]+tab[3,3]+tab[4,4])/sum(tab)
list(Matrix_Uji=tab, Akurasi=accuracy, Total_data=sum(tab))
print(scimago_ctree)
plot(scimago_ctree)
testPred <- predict(scimago_ctree, newdata = testData)
tab2 <- table(testPred, testData$SJR_Best_Quartile)
accuracy2 <-
sum(tab2[1,1]+tab2[2,2]+tab2[3,3]+tab2[4,4])/sum(tab2)
list(Matrix_Uji=tab2, Akurasi=accuracy2, Total_data=sum(tab2))

library(data.tree)
library(rpart)
library(yaml)
x_trainData <- scimago[,1:9]
y_trainData <- scimago[,10]
x <- cbind(x_trainData, y_trainData)
fit <- rpart(y_trainData ~ ., data=x, method="class")
summary(fit)
SJR = c(0.1)
H_index = c(39)
x_test = data.frame(SJR, H_index)
predicted = predict(fit, x_test)
predicted
```

### Lampiran 3. Hasil Klasifikasi

```
> data <- read.delim("clipboard")
> scimago <- data
> str(scimago)
'data.frame': 23929 obs. of 10 variables:
 $ SJR : num 39.3 33.2 29.7 29.7 27.6 ...
 $ SJR_Best_Quartile : Factor w/ 4 levels "Q1","Q2","Q3",...: 1
1 1 1 1 1 1 1 1 1 ...
 $ H_index : int 131 292 316 352 267 655 205 271 151
355 ...
 $ Total_Docs_2016 : int 43 166 146 152 23 693 40 46 18 115
...
 $ Total_Docs_3years : int 141 615 581 535 72 1885 121 123 43
513 ...
 $ Total_Refs : int 3503 8029 7719 9128 4155 29440 2401
12278 4779 9863 ...
 $ Total_Cites_3years : int 11929 7131 8256 8150 2513 42666 986
4383 1512 7806 ...
 $ Citable_Docs_3years: int 118 183 195 214 72 1690 120 117 42
207 ...
 $ Cites_Doc_2years : num 128.8 39.7 36.5 45.1 28.8 ...
 $ Ref_Doc : num 81.5 48.4 52.9 60 180.7 ...
> set.seed(1234)
> ind <- sample(2, nrow(scimago), replace=TRUE, prob=c(0.7,
0.3))
> trainData <- scimago[ind==1,]
> testData <- scimago[ind==2,]
> install.packages("party")
Installing package into 'C:/Users/Dell/Documents/R/win-
library/3.4'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
also installing the dependencies 'TH.data', 'multcomp',
'mvtnorm', 'modeltools', 'strucchange', 'coin', 'sandwich'

trying URL 'https://cloud.r-
project.org/bin/windows/contrib/3.4/TH.data_1.0-8.zip'
Content type 'application/zip' length 5213284 bytes (5.0 MB)
downloaded 5.0 MB

trying URL 'https://cloud.r-
project.org/bin/windows/contrib/3.4/multcomp_1.4-8.zip'
Content type 'application/zip' length 633366 bytes (618 KB)
downloaded 618 KB

trying URL 'https://cloud.r-
project.org/bin/windows/contrib/3.4/mvtnorm_1.0-7.zip'
Content type 'application/zip' length 233555 bytes (228 KB)
downloaded 228 KB

trying URL 'https://cloud.r-
project.org/bin/windows/contrib/3.4/modeltools_0.2-21.zip'
Content type 'application/zip' length 138817 bytes (135 KB)
downloaded 135 KB
```

```

trying URL 'https://cloud.r-project.org/bin/windows/contrib/3.4/strucchange_1.5-1.zip'
Content type 'application/zip' length 853809 bytes (833 KB)
downloaded 833 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/3.4/coin_1.2-2.zip'
Content type 'application/zip' length 1461439 bytes (1.4 MB)
downloaded 1.4 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/3.4/sandwich_2.4-0.zip'
Content type 'application/zip' length 1243236 bytes (1.2 MB)
downloaded 1.2 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/3.4/party_1.2-4.zip'
Content type 'application/zip' length 720029 bytes (703 KB)
downloaded 703 KB

package 'TH.data' successfully unpacked and MD5 sums checked
package 'multcomp' successfully unpacked and MD5 sums checked
package 'mvtnorm' successfully unpacked and MD5 sums checked
package 'modeltools' successfully unpacked and MD5 sums checked
package 'strucchange' successfully unpacked and MD5 sums checked
package 'coin' successfully unpacked and MD5 sums checked
package 'sandwich' successfully unpacked and MD5 sums checked
package 'party' successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\Dell\AppData\Local\Temp\RtmpkJyziJ\downloaded_packages
> library(party)
Loading required package: grid
Loading required package: mvtnorm
Loading required package: modeltools
Loading required package: stats4
Loading required package: strucchange
Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

  as.Date, as.Date.numeric

Loading required package: sandwich
Warning messages:
1: package 'party' was built under R version 3.4.4
2: package 'strucchange' was built under R version 3.4.4
3: package 'sandwich' was built under R version 3.4.4
> myFormula <- SJR_Best_Quartile ~ SJR + H_index +
Total_Docs_2016 + Total_Docs_3years + Total_Refs +
Total_Cites_3years + Citable_Docs_3years + Cites_Doc_2years +
Ref_Doc
> scimago_ctree <- ctree(myFormula, data=trainData)

```

```

> tab <- table(predict(scimago_ctree),
trainData$SJR_Best_Quartile)
> accuracy <- sum(tab[1,1]+tab[2,2]+tab[3,3]+tab[4,4])/sum(tab)
> list(Matrix_Uji=tab, Akurasi=accuracy, Total_data=sum(tab))
$Matrix_Uji

      Q1  Q2  Q3  Q4
Q1 3953 817  42   0
Q2  535 2474 891  41
Q3  229  551 2182 529
Q4   54  352  873 3211

$Akurasi
[1] 0.7063464

$Total_data
[1] 16734

> print(scimago_ctree)

      Conditional inference tree with 66 terminal nodes

Response:  SJR_Best_Quartile
Inputs:  SJR, H_index, Total_Docs_2016, Total_Docs_3years,
Total_Refs, Total_Cites_3years, Citable_Docs_3years,
Cites_Doc_2years, Ref_Doc
Number of observations:  16734

1) H_index <= 39; criterion = 1, statistic = 4810.206
  2) SJR <= 0.16; criterion = 1, statistic = 3675.276
    3) SJR <= 0.129; criterion = 1, statistic = 654.826
      4) SJR <= 0.108; criterion = 1, statistic = 242.03
        5) SJR <= 0.1; criterion = 1, statistic = 45.532
          6)* weights = 614
            5) SJR > 0.1
              7) H_index <= 3; criterion = 1, statistic = 52.188
                8) Total_Docs_3years <= 23; criterion = 0.996,
                  statistic = 15.465
                  9) SJR <= 0.106; criterion = 1, statistic = 20.01
                    10) SJR <= 0.105; criterion = 0.995, statistic =
14.982
                      11)* weights = 272
                        10) SJR > 0.105
                          12)* weights = 19
                            9) SJR > 0.106
                              13)* weights = 21
                                8) Total_Docs_3years > 23
                                  14)* weights = 436
                                    7) H_index > 3
                                      15) SJR <= 0.105; criterion = 1, statistic = 27.777
                                        16) SJR <= 0.103; criterion = 0.983, statistic =
9.595
                                          17) H_index <= 9; criterion = 0.976, statistic =
8.974
                                            18)* weights = 331
                                              17) H_index > 9

```

```

19) Ref_Doc <= 41; criterion = 0.997,
statistic = 12.79
    20)* weights = 66
    19) Ref_Doc > 41
    21)* weights = 8
    16) SJR > 0.103
    22)* weights = 114
    15) SJR > 0.105
    23)* weights = 151
4) SJR > 0.108
    24) Ref_Doc <= 37.82; criterion = 1, statistic = 85.161
    25) H_index <= 3; criterion = 1, statistic = 53.594
    26)* weights = 449
    25) H_index > 3
    27) Ref_Doc <= 24.2; criterion = 1, statistic =
26.536
    28)* weights = 577
    27) Ref_Doc > 24.2
    29)* weights = 200
    24) Ref_Doc > 37.82
    30)* weights = 282
3) SJR > 0.129
    31) Ref_Doc <= 55.06; criterion = 1, statistic = 116.137
    32) Ref_Doc <= 27.03; criterion = 1, statistic = 52.39
    33) H_index <= 3; criterion = 1, statistic = 27.65
    34)* weights = 115
    33) H_index > 3
    35) Ref_Doc <= 7.45; criterion = 0.988, statistic =
15.593
    36)* weights = 226
    35) Ref_Doc > 7.45
    37)* weights = 350
    32) Ref_Doc > 27.03
    38) Cites_Doc_2years <= 0.21; criterion = 1, statistic
= 39.2
    39)* weights = 252
    38) Cites_Doc_2years > 0.21
    40)* weights = 218
    31) Ref_Doc > 55.06
    41)* weights = 105
2) SJR > 0.16
    42) SJR <= 0.657; criterion = 1, statistic = 1710.478
    43) SJR <= 0.293; criterion = 1, statistic = 2060.564
    44) SJR <= 0.201; criterion = 1, statistic = 351.258
    45) Ref_Doc <= 39.47; criterion = 1, statistic =
90.365
    46) SJR <= 0.182; criterion = 0.998, statistic =
18.998
    47) Ref_Doc <= 33.77; criterion = 0.95, statistic
= 12.577
    48)* weights = 372
    47) Ref_Doc > 33.77
    49)* weights = 51
    46) SJR > 0.182
    50)* weights = 450
    45) Ref_Doc > 39.47

```

```

51)* weights = 272
44) SJR > 0.201
52) SJR <= 0.245; criterion = 1, statistic = 135.844
53) Ref_Doc <= 44.57; criterion = 1, statistic =
51.283
54) SJR <= 0.226; criterion = 0.995, statistic =
17.724
55) H_index <= 18; criterion = 0.979, statistic
= 14.477
56)* weights = 442
55) H_index > 18
57)* weights = 116
54) SJR > 0.226
58)* weights = 369
53) Ref_Doc > 44.57
59) Total_Cites_3years <= 95; criterion = 1,
statistic = 24.044
60)* weights = 176
59) Total_Cites_3years > 95
61)* weights = 7
52) SJR > 0.245
62) Cites_Doc_2years <= 0.3; criterion = 1,
statistic = 34.185
63)* weights = 131
62) Cites_Doc_2years > 0.3
64) Ref_Doc <= 33.94; criterion = 1, statistic =
23.405
65) SJR <= 0.268; criterion = 0.995, statistic =
17.451
66)* weights = 260
65) SJR > 0.268
67)* weights = 263
64) Ref_Doc > 33.94
68) Total_Docs_2016 <= 112; criterion = 0.999,
statistic = 22.071
69)* weights = 243
68) Total_Docs_2016 > 112
70)* weights = 20
43) SJR > 0.293
71) SJR <= 0.391; criterion = 1, statistic = 411.525
72) Total_Docs_2016 <= 80; criterion = 1, statistic =
37.21
73) Total_Cites_3years <= 50; criterion = 1,
statistic = 23.391
74) Total_Refs <= 1670; criterion = 1, statistic =
24.081
75) Total_Refs <= 478; criterion = 0.965,
statistic = 13.371
76)* weights = 140
75) Total_Refs > 478
77)* weights = 273
74) Total_Refs > 1670
78) Cites_Doc_2years <= 0.63; criterion = 0.953,
statistic = 10.446
79)* weights = 19
78) Cites_Doc_2years > 0.63

```

```

      80)* weights = 8
      73) Total_Cites_3years > 50
      81)* weights = 642
      72) Total_Docs_2016 > 80
      82) Total_Docs_2016 <= 629; criterion = 1, statistic
= 24.187
      83) SJR <= 0.31; criterion = 0.981, statistic =
14.619
      84)* weights = 57
      83) SJR > 0.31
      85)* weights = 215
      82) Total_Docs_2016 > 629
      86)* weights = 7
      71) SJR > 0.391
      87) SJR <= 0.585; criterion = 1, statistic = 126.845
      88) SJR <= 0.49; criterion = 1, statistic = 42.215
      89) Ref_Doc <= 37.53; criterion = 0.991, statistic
= 16.293
      90)* weights = 583
      89) Ref_Doc > 37.53
      91)* weights = 325
      88) SJR > 0.49
      92) Cites_Doc_2years <= 1; criterion = 1,
statistic = 29.922
      93) Ref_Doc <= 26.17; criterion = 0.956,
statistic = 10.588
      94)* weights = 83
      93) Ref_Doc > 26.17
      95) SJR <= 0.524; criterion = 0.956, statistic
= 10.592
      96)* weights = 73
      95) SJR > 0.524
      97)* weights = 94
      92) Cites_Doc_2years > 1
      98)* weights = 346
      87) SJR > 0.585
      99) Total_Docs_2016 <= 14; criterion = 0.999,
statistic = 18.496
      100)* weights = 44
      99) Total_Docs_2016 > 14
      101) Total_Docs_2016 <= 72; criterion = 1,
statistic = 24.974
      102) Ref_Doc <= 30.82; criterion = 0.972,
statistic = 11.539
      103)* weights = 53
      102) Ref_Doc > 30.82
      104)* weights = 158
      101) Total_Docs_2016 > 72
      105)* weights = 84
      42) SJR > 0.657
      106) SJR <= 0.834; criterion = 1, statistic = 56.764
      107) SJR <= 0.726; criterion = 0.989, statistic = 13.436
      108)* weights = 239
      107) SJR > 0.726
      109)* weights = 289
      106) SJR > 0.834

```

```

110) SJR <= 1.325; criterion = 0.998, statistic = 16.728
    111)* weights = 576
110) SJR > 1.325
    112)* weights = 329
1) H_index > 39
  113) SJR <= 0.704; criterion = 1, statistic = 349.351
    114) SJR <= 0.191; criterion = 1, statistic = 301.538
      115)* weights = 15
    114) SJR > 0.191
      116) SJR <= 0.394; criterion = 1, statistic = 248.801
      117) SJR <= 0.254; criterion = 0.957, statistic = 12.908
      118)* weights = 28
      117) SJR > 0.254
      119)* weights = 147
    116) SJR > 0.394
      120) SJR <= 0.638; criterion = 1, statistic = 105.07
      121) SJR <= 0.546; criterion = 1, statistic = 33.793
      122)* weights = 333
      121) SJR > 0.546
      123)* weights = 281
      120) SJR > 0.638
      124)* weights = 207
    113) SJR > 0.704
      125) SJR <= 1.039; criterion = 1, statistic = 86.081
      126) SJR <= 0.833; criterion = 1, statistic = 20.15
      127)* weights = 412
      126) SJR > 0.833
      128)* weights = 542
      125) SJR > 1.039
      129) SJR <= 1.422; criterion = 1, statistic = 19.648
      130)* weights = 755
      129) SJR > 1.422
      131)* weights = 1399
> plot(scimago_ctree)
> testPred <- predict(scimago_ctree, newdata = testData)
> tab2 <- table(testPred, testData$SJR_Best_Quartile)
> accuracy2 <-
sum(tab2[1,1]+tab2[2,2]+tab2[3,3]+tab2[4,4])/sum(tab2)
> list(Matrix_Uji=tab2, Akurasi=accuracy2, Total_data=sum(tab2))
$Matrix_Uji

testPred   Q1    Q2    Q3    Q4
   Q1 1636  351   20    0
   Q2  241 1093  328   20
   Q3   92  250  996  237
   Q4   17  144  389 1381

$Akurasi
[1] 0.7096595

$Total_data
[1] 7195

```



