

**BAB II**  
**TINJAUAN PUSTAKA**

**2.1 Penelitian Terkait**

No	Penelitian	Variabel	Hasil
1.	Nadia Tri Hantari (2018)	Metode <i>Naïve Bayes</i> dan Algoritma C4.5, analisis kelulusan	Mengetahui hasil keakurasian kedua metode mana yang lebih baik dalam menganalisis kelulusan Mahasiswa dan hasilnya Algoritma C4.5 memiliki tingkat akurasi 85.8156% sedangkan <i>Naïve Bayes</i> memiliki keakuratan 79.4326%.
2.	Mongan, Arie, Agustinus (2017)	Algoritma <i>Naive Bayes</i> , prediksi masa studi	Dapat dilakukan prediksi terhadap ketepatan waktu studi mahasiswa, hasil pengujian didapat nilai akurasi rata-rata sebesar 85.17% sedangkan nilai akurasi tertinggi sebesar 88.96%.
3.	Bustami (2014)	Algoritma <i>Naïve Bayes</i> , mengklasifikasi nasabah	Untuk memperkirakan nasabah yang bergabung sehingga perusahaan bisa mengambil keputusan menerima atau menolak calon nasabah tersebut.
4.	Ferian Anggoro H (2017)	Metode <i>Naïve Bayes Classifier</i> , penerimaan beasiswa	Untuk melakukan pengambilan keputusan dalam menentukan diterima atau ditolaknya mahasiswa yang mendaftarkan beasiswa.
5.	Muchamad Hasan B (2015)	Algoritma <i>Naïve Bayes</i> ,	Dapat meminimalisir dampak dari kekeliruan dalam proses penjurusan siswa dan menggunakan tools matlab

		prediksi penjurusan	
--	--	------------------------	--

**Tabel 2.1 Penelitian Terdahulu**

## 2.2 Prediksi

Prediksi menurut kamus besar bahasa Indonesia, prediksi adalah hasil dari kegiatan memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan inout bagi proses perencanaan dan pengambilan keputusan. Tujuan dari adanya prediksi tersebut adalah untuk memperoleh informasi mengenai perubahan dimasa yang akan datang yang akan mempengaruhi terhadap implementasi kebijakan serta konsekuensinya. Didalam memprediksikan kebutuhan yang akan datang dengan berpijak pada masa lalu, dibutuhkan seseorang yang memiliki daya sensitifitas tinggi dan mampu membaca kemungkinan- kemungkinan dimasa yang akan datang . (Mongan, Arie, Agustinus 2017)

## 2.3 Perguruan Tinggi

Perguruan tinggi adalah satuan pendidikan yang menyelenggarakan pendidikan tinggi, yang kelembagaannya dapat berupa akademi, politeknik, sekolah tinggi, institusi atau universitas. Pendidikan tinggi adalah jenjang pendidikan formal setelah pendidikan menengah. Pendidikan tinggi menyelenggarakan kegiatan pembelajaran, penelitian, dan pengabdian kepada masyarakat dengan menjunjung tinggi nilai-nilai persatuan bangsa sehingga mampu menghasilkan lulusan yang kompeten(natasya dkk 2020).

Perguruan tinggi menurut Nano Supriono adalah satuan pendidikan yang padanya diselenggarakan jenjang pendidikan tinggi dimana peserta didiknya

disebut mahasiswa, sedangkan tenaga pendidiknya disebut Dosen. Perguruan tinggi adalah satuan pendidikan yang menyelenggarakan pendidikan tinggi. Hal tersebut sebagaimana menurut Undang-undang No. 12 Tahun 2012 Tentang Pendidikan Tinggi dalam Pasal 1 Ayat 6. Dengan demikian maka perguruan tinggi merupakan jenjang pendidikan tinggi, yaitu setingkat di atas jenjang pendidikan dasar dan pendidikan menengah. Penjelasan tersebut searah dengan keterangan dalam Undang-undang RI No. 20 tahun 2003 tentang Sisdiknas bab VI bagian keempat tentang pendidikan tinggi pada pasal 19 nomor 1 djelaskan “pendidikan tinggi merupakan jenjang pendidikan setelah pendidikan menengah yang mencakup program pendidikan diploma, sarjana, magister, spesialis, dan doctor yang diselenggarakan perguruan tinggi”(Arifki 2015).

#### **2.4 Dunia Kerja**

Pengertian Dunia Kerja adalah gambaran tentang beberapa jenis dan proporsi pekerjaan yang ada seperti dalam bidang pertanian, usaha dan perkantoran, rekayasa, kesehatan, militer kemasyarakatan, kerumah tanggaan, dan seni budaya. Dalam era globalisasi seluruh dunia kerja dan industri berusaha meningkatkan efisiensi dan produktifitas kerja. Adanya peningkatan efisiensi dan produktifitas kerja menunjukkan bahwa perusahaan telah melaksanakan re-engineering dan re-strukturings dalam rangka mempersingkat proses produksi.(Nouval, 2017)

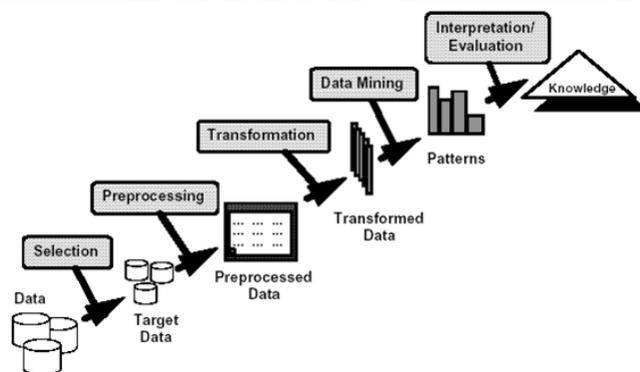
Kesiapan untuk memasuki dunia kerja ada beberapa aspek yang harus di siapkan yaitu:

1. kepercayaan diri, yaitu mempunyai kepercayaan diri yang tinggi dengan bekal pengetahuan dan keterampilan yang dimiliki dapat menyesuaikan diri dengan lingkungan kerja,
2. komitmen, yaitu kemauan/kesungguhan dalam melaksanakan pekerjaan sesuai dengan aturan yang berlaku,
3. inisiatif/kreatif, yaitu mempunyai inisiatif dan kreatifitas yang tinggi dalam mengembangkan suatu keputusan tentang tugas yang di berikan,
4. ketekunan dalam bekerja, yaitu mempunyai keyakinan dan kesabaran dalam menyelesaikan pekerjaan,

5. kecakapan kerja, yaitu mempunyai kemampuan yang tinggi dalam melaksanakan pekerjaan baik dari segi pengetahuan, maupun keterampilan,
6. kedisiplinan, yaitu mempunyai sikap disiplin yang tinggi, patuh dan taat mengikuti segala peraturan dan ketentuan yang berlaku,
7. motivasi berprestasi, yaitu mempunyai kemauan yang tinggi untuk mengembangkan diri,
8. kemampuan bekerja sama, yaitu mempunyai sikap terbuka dan siap untuk bekerja sama dengan siapa saja dan bekerja dalam satu tim,
9. tanggung jawab, yaitu mempunyai rasa tanggung jawab yang tinggi terhadap pekerjaan yang diberikan,
10. kemampuan berkomunikasi, yaitu mempunyai kemampuan berkomunikasi dengan baik, seperti penguasaan bahasa teknik, bahasa asing dan lain-lain.

## 2.5 Data Mining

Data mining merupakan salah satu bagian atau proses utama dari *Knowledge Discovery in Database (KDD)* yang berbentuk kegiatannya yaitu mengumpulkan dan menggunakan data masa lalu untuk menemukan keterturan, pola atau hubungan dalam suatu set data yang lebih besar . secara garis besar KDD meliputi tiga Tahapan, yaitu *pre-processing*, *process (data mining)* dan *post processing*. Kesimpulannya, data mining adalah proses penemuan pola-pola tertentu dari sebuah data atau basis data yang berukuran besar untuk memperoleh informasi yang sangat berguna (Erwin.2017) .



## Gambar 2.1 tahapan KDD

Tahapan–tahapan KDD ( *Knowledge Discovery in Database*) (kusrini ,  
Emha. 2009)

### 1. Data Selesction

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahapan penggalian informasi dalam KDD dimulai.

### 2. Pre-processing/cleaning

Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

### 3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.

### 4. Data Mining

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik dan metode tertentu.

### 5. Interpretation/Evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

Teknik data mining dapat menyelesaikan beberapa masalah yang sesuai dengan ciri sebagai berikut (Erwin.2017) :

1. Memerlukan keputusan yang bersifat berdasarkan pengetahuan
2. Berada pada lingkungan yang dinamis
3. Metode yang ada bersifat sub-optimal
4. Data yang dapat diakses tersebut, cukup dan relevan
5. Memberikan keuntungan yang tinggi apabila disertai keputusan yang tepat

Berikut pengelompokan data mining berdasarkan fungsi dan tujuan (Efori.2020) :

1. Deskripsi

Proes dekripsi bertujuan untuk menemukan/ identifikasi pola yang sering muncul dan mengubah pola tersebut menjadi aturan yang dapat dipergunakan untuk mempermudah suatu aktivitas.

2. Klasifikasi

Pengelompokan berdasarkan hubungan antara variabel kriteria dengan variabel target.

3. Prediksi

Fungsi data mining yang sering digunakan adalah untuk memprediksi. Nilai dari hasil prediksi akan digunakan dimasa yang akan datang berdasarkan data-data sebelumnya.

4. Estimasi

Estimasi adalah perkiraan/ prediksi, sehingga estimasi hampir sama dengan klasifikasi, perbedaan terletak pada bentuk pengelompokan, dimana estimasi pengelompokan kearah numerik dan bukan kearah kategori.

5. Pengklasteran

Pengklasteran merupakan pengelompokan data yang memiliki kemiripan nilai (homogen ).

6. Asosiasi

Asosiasi adalah kumpulan, himpunan, persatuan atau persekutuan.

## 2.6 Naïve Bayes

Naïve Bayes merupakan sebuah pengklasifikasian probabilistic sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling bergantung yang diberikan oleh nilai pada variable kelas.

Definisi lain mengatakan Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistic yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya(Dinda, Maulana, Muhammad 2020)

Algoritma Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk *machine learning dan data mining*. Performa naïve bayes yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keindependenan atribut (tidak ada kaitan antara atribut).

Karakteristik Naïve Bayes : klasifikasi dengan Naïve Bayes bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas.

Keunggulan dan kelemahan Naïve Bayes (Yusniar, Nuraini, Noviana 2020)

1. Keunggulan
    - a. Mudah diimplementasikan
    - b. Memberikan hasil yang baik untuk banyak kasus
    - c. Membutuhkan jumlah data yang sedikit untuk klasifikasi
    - d. Cepat, mudah dan efisien untuk dibuat
    - e. Kokoh pada atribut Yng tidak relevan
  2. Kelemahan
    - a. Harus mengasumsi bahwa antar fitur tidak terkait (independet). Dalam realita, keterkaitan itu ada. Keterkaitan tersebut tidak dapat dimodelkan oleh Naïve Bayes
    - b. Independen antar tribut membuat akurasi menjadi berkurang
    - c. Tidak berlaku apabila nilai probabilitasnya adalah 0
- Persamaan dari terorema Naïve Bayes (Cahya,2018)

$$P(S|X) = \frac{P(X|Y).P(Y)}{P(X)} \dots\dots\dots(1)$$

Keterangan :

X = data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (posterior probability)

P(H) = probabilitas hipotesis H (prior probability )

P(X|H) = probabilitas X berdasarkan kondisi tersebut

P(X) = probabilitas dari X

## 2.7 Confusin matrix

Confusion matrix adalah suatu metode atau cara yang dapat digunakan untuk melakukan perhitungan dan pengolahan akurasi pada suatu konsep data mining atau sitem pendukung keputusan. Dalam proses pengukuran kinerja menggunakan confusion matrix, terdapat 4 istilah atau sebutan sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah (Hantari, 2018).

1. TP (True Positive) adalah jumlah data yang positif dinilai benar
2. TN (True Negative) adalah jumlah data negative yang terdeteksi dengan benar
3. FP (False Positif) adalah data negative tapi terdeteksi bernilai positif
4. FN (False Negatif) adalah data positif tapi terdeteksi dengan negative (salah)

Akurasi merupakan presentase dari seluruh jumlah data yang teridentifikasi dan ternilai. Aturan akurasi antara lain :

$$\text{Akurasi} = ((\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN})) * 100\% \dots\dots\dots(2)$$

## 2.8 Bahasa R

R adalah nama bahasa pemrograman komputer yang ditujukan secara khusus untuk menangani komputasi statistik dan memudahkan penyajian grafik. Bahasa ini diciptakan oleh Ross Ihaka dan Robert Gentleman di Universitas Auckland, Selandia Baru. Versi pertama dirilis pada tahun 1995. Adapun nama R disematkan berdasarkan nama depan kedua penciptanya. R bisa dipakai menggunakan lisensi GNU GPL, selain itu tersedia versi binernya untuk sistem operasi Linux, Mac OS X, dan Windows (Abdul, 2019).

Keutamaan bahasa R (AR Effendie 2018) :

1. Implementasi domain publik dari bahasa statistik S yang dianggap luas.
2. Dalam banyak hal, dapat dibandingkan, dan seringkali lebih unggul, dalam kekuatan untuk produk komersial
3. Tersedia untuk Linux, Mac OS X, dan Windows
4. Selain mengaktifkan operasi statistik, ini adalah bahasa pemrograman umum, jadi bisa mengotomatiskan analisis dan membuat fungsi baru

5. Struktur pemrograman yang berorientasi pada objek dan fungsional
6. Dataset disimpan diantara sesi, jadi tidak perlu memuat ulang setiap kali
7. Bentuk perangkat lunak terbuka

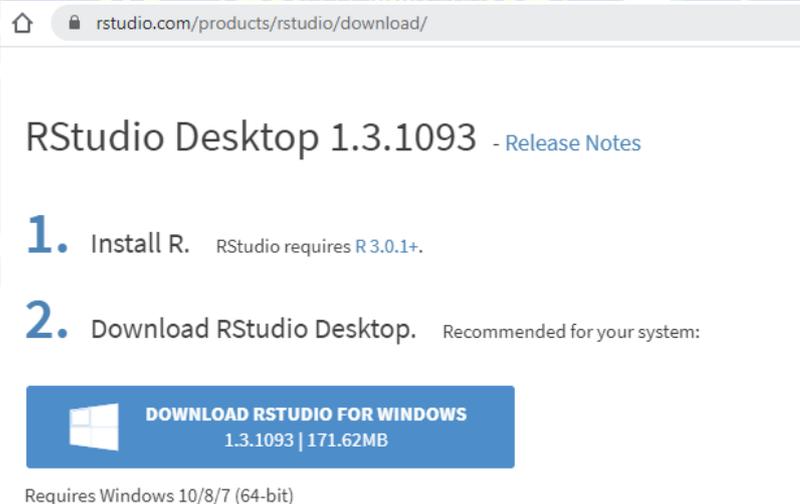
Pada penelitian ini menggunakan RStudio versi 1.3.1093 dan *R for Windows* versi 4.0.3. Untuk instalasi RStudio di *Windows* yang pertama ialah download Bahasa R, buka website [www.r-project.org](http://www.r-project.org) kemudian pilih link pengunduhan sesuai dengan system operasi yang digunakan.



**Gambar 2.2** Download R

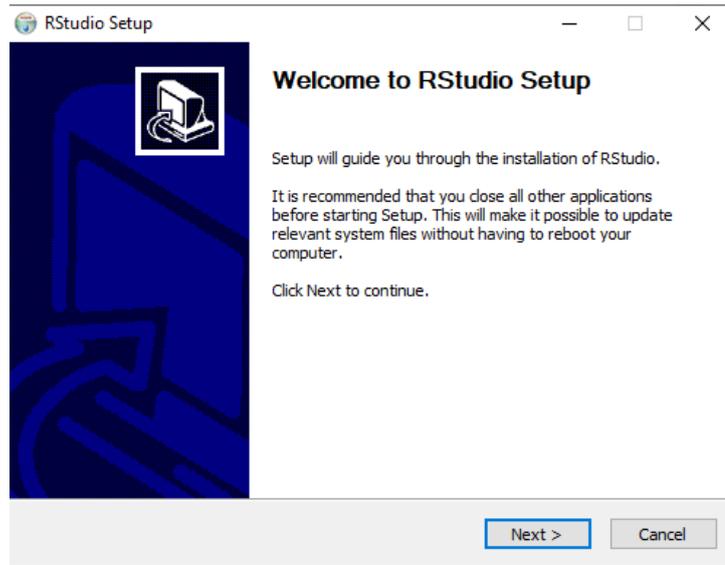
Setelah proses download selesai jalankan bahasa R yang sudah didownload sebelumnya, untuk proses penginstalan cukup klik next sampai finish.

Kemudian download dan install RStudio, buka alamat [www.rstudio.com](http://www.rstudio.com) kemudian pilih RStudio Desktop dan pilih yang Open Source License untuk memilih yang Free.



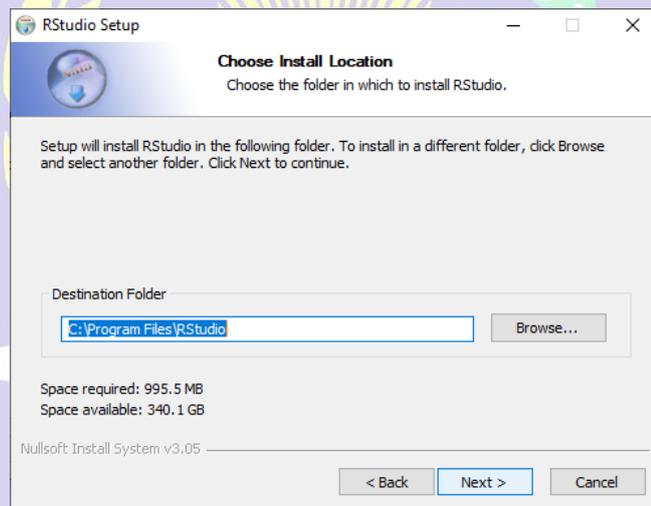
**Gambar 2.3** Download RStudio

Jika proses download sudah selesai, Langkah selanjutnya yaitu install RStudio. Langkah pertama klik next



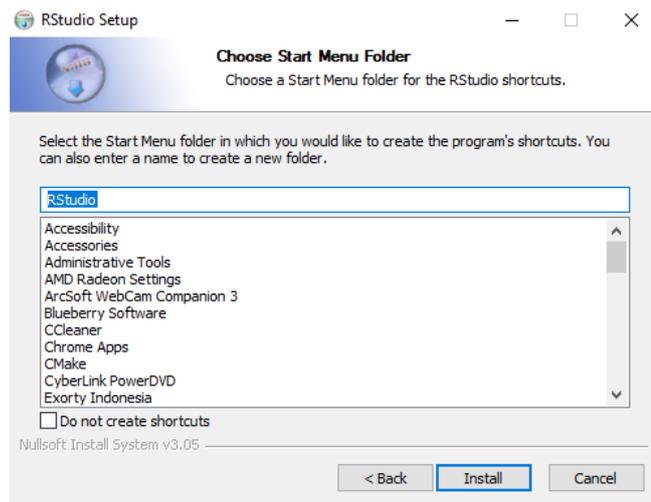
**Gambar 2.4** Install Rstudio langkah pertama

Langkah kedua, untuk penyimpanan disini menggunakan data C, kemudian klik next



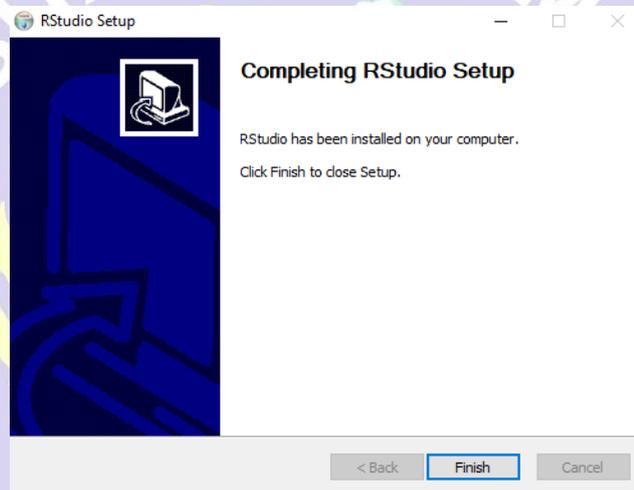
**Gambar 2.5** Install Rstudio langkah kedua

Langkah ketiga, memilih shortcut setelah selesai instalasi, kemudian klik next



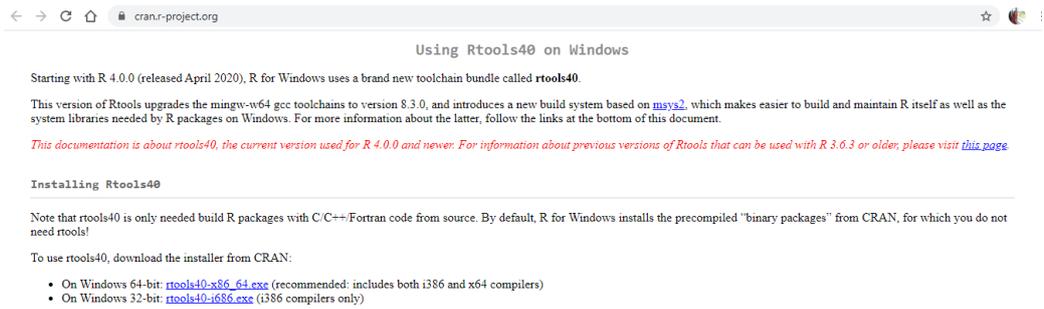
**Gambar 2.6** Install RStudio langkah ketiga

Langkah keempat klik finish menandakan instalasi sudah selesai



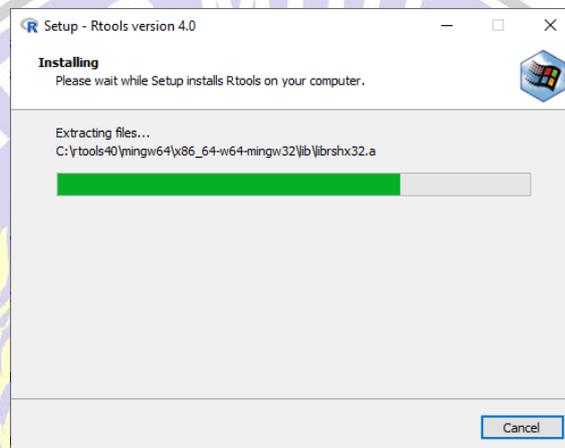
**Gambar 2.7** Install RStudio langkah keempat

Setelah selesai proses instalasi RStudio, kemudian install RTools yang akan digunakan untuk proses analisis. Buka alamat <https://cran.r-project.org/bin/windows/Rtools/> , kemudian pilih CRAN sesuai bit pada windows.



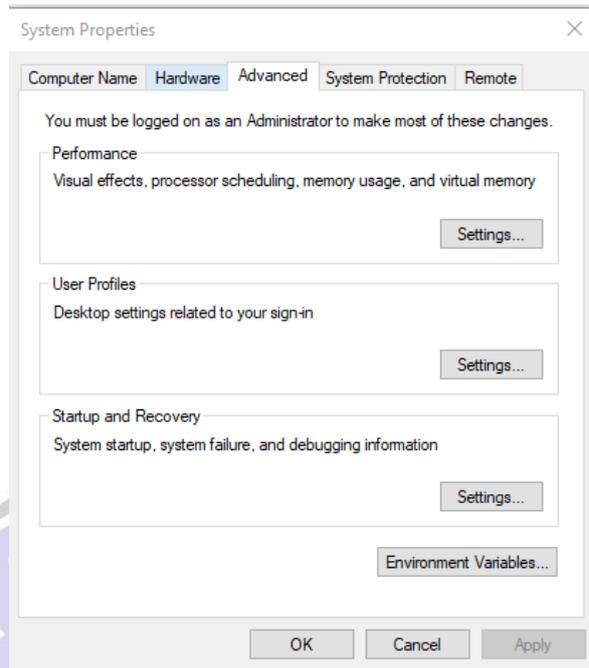
**Gambar 2.8** Download RTools

Setelah selesai proses download, kemudian klik file RTools yang sudah didownload sebelumnya tunggu hingga proses selesai.



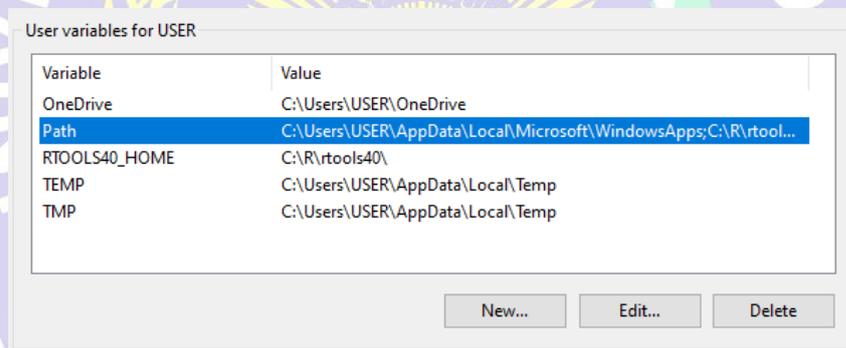
**Gambar 2.9** Install RTools

Setelah install selesai kemudian buka Edit the system environment variables pilih advanced kemudian pilih Environment Variables



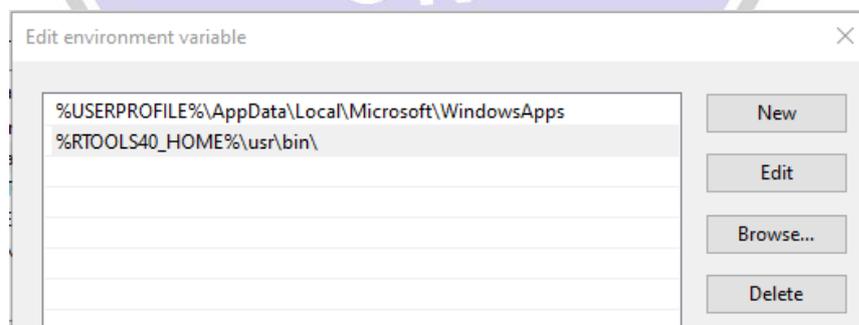
**Gambar 2.10** Edit the system

Kemudian pada “User variables for USER” klik dua kali Path



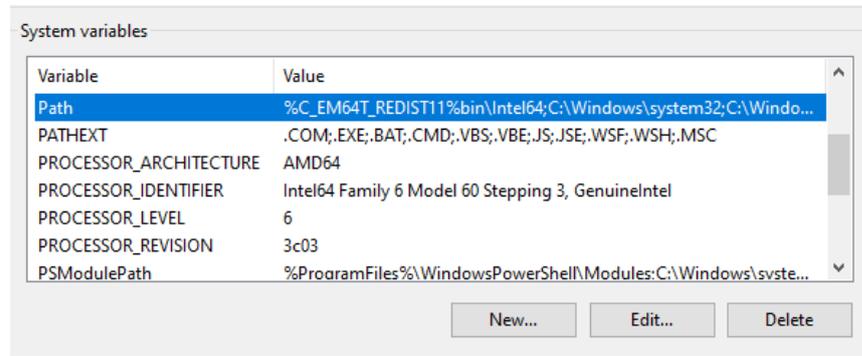
**Gambar 2.11** Edit Path User variables for USER

Klik New kemudian ketik “%RTOOLS40\_HOME%\usr\bin\



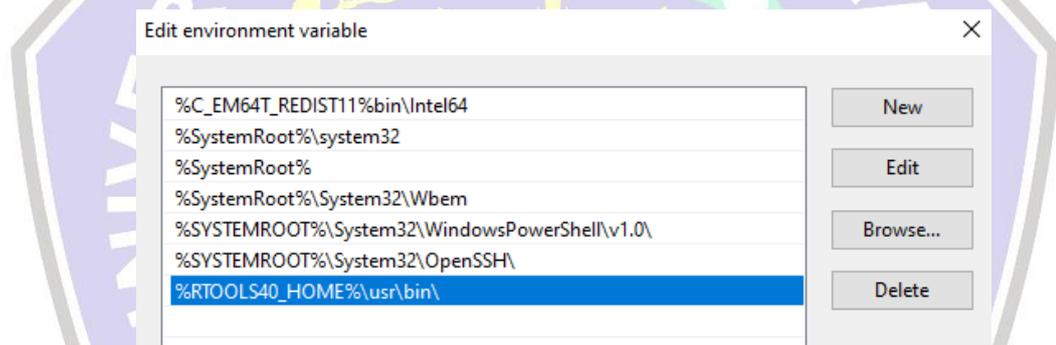
**Gambar 2.12** menambah Path User variables for USER

Pada System variables juga klik dua kali pada Path



**Gambar 2.13** menambah Path System variables

Kemudian klik New dan ketik “%RTOOLS40\_HOME%\usr\bin\”



**Gambar 2.14** menambah Path System variables2

Setelah proses selesai kemudian restart RStudio dan buka kembali sehingga bisa digunakan untuk proses instalasi.