

BAB II LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian Berjudul **“CRICKET SCRAPPER: A TOOL DEVELOPED TO EXTRACT CRICKET PLAYERS DATA”** dilakukan oleh (Asif et al., 2019) menjelaskan mengenai kegunaan dari teknologi *web scraping* yang dapat membantu mengumpulkan informasi dari mengekstrak konten dari website-website agar orang dapat memperoleh informasi dengan mudah. Program yang dihasilkan dari penelitian menggunakan Bahasa pemrograman *C#* dan *library Anglesharp*.

Selanjutnya, penelitian berjudul **“EVALUATION OF HAP, ANGLESHARP AND HTMLDOCUMENT IN WEB CONTENT EXTRACTION”** dilakukan oleh (Uzun, Buluş, Doruk, & Özhan, 2017) penelitian ini membahas tentang hasil dari evaluasi penggunaan *library* seperti *HAP*, *Anglesharp*, *HTMLDocument* untuk kegiatan mengekstraksi konten website.

Kemudian hasil dari evaluasi tersebut menunjukkan bahwa *Anglesharp* adalah *library* yang lebih baik diantara kedua *library* tersebut.

Kemudian penelitian lainnya yang berjudul **“THE DEVELOPMENT OF INTEGRATED BALI TOURISM INFORMATION PORTAL USING WEB SCRAPPING AND CLUSTERING METHODS”** dilakukan oleh (Darmawiguna, Pradnyana, & Santyadiputra, 2019), penelitian ini menjelaskan mengenai pengertian dari *web scrapping*, dan kemudian juga berisikan tentang bagaimana teknologi *web scrapping* serta metode yang digunakan dapat membantu mengembangkan wisata-wisata yang ada di Bali dengan cara menghimpun informasi tentang tempat wisata tersebut.

Dapat dilihat juga penelitian berjudul **“IMPLEMENTASI WEB SCRAPING DAN TEXT MINING UNTUK AKUISISI DAN KATEGORISASI INFORMASI LAMAN WEB TENTANG HIDROPONIK”**

dilakukan oleh (Priyanto & Ma’arif, 2018) Penelitian ini bertujuan untuk mengumpulkan beberapa informasi dari beberapa website mengenai hidroponik, dimana informasi yang dikumpulkan akan dikelompokkan kebeberapa kategori secara otomatis. Aplikasi dikembangkan menggunakan metode penelitian *waterfall* yang memiliki tahapan sebagai berikut, identifikasi dan analisis kebutuhan, desain sistem, implementasi sistem, dan pengujian sistem.

Dalam penelitian berikut berjudul **“IMPLEMENTASI LEVENSHTTEIN DISTANCE PADA APLIKASI PENCARIAN BARANG DI BERBAGAI E-MARKETPLACE MENGGUNAKAN TEKNIK WEB SCRAPING”**

dilakukan oleh (Arsyad, Pramono, Isnawaty, Yamin, & Ihsan, 2019). Penelitian ini bertujuan untuk mengumpulkan beberapa informasi dari beberapa website mengenai hidroponik, dimana informasi yang dikumpulkan akan dikelompokkan kebeberapa kategori secara otomatis. Aplikasi dikembangkan menggunakan metode penelitian *waterfall* yang memiliki tahapan sebagai berikut, identifikasi dan menganalisis tujuan yang ingin dicapai, mendesain sistem, diteruskan pengimplementasian sistem, dan terakhir menguji sistem.

Tabel 2.1 Hasil Penelitian

Nama	Tahun	Kesimpulan
Asif, Haque, Hina, Qaim Haider, & Khan	2019	Penulis menjelaskan teknologi <i>web scraping</i> yang berguna untuk pengumpulan data yang diolah menjadi informasi yang berguna..
Uzun, Bulus, Doruk, & Ozhan	2017	Hasil penelitian menerangkan bahwa <i>library Anglesharp</i> adalah

		<i>library web scraping</i> terbaik diantara <i>HAP</i> , dan <i>HTMLDocument</i> .
Darmawiguna, Pradnyana, & Santyadiputra	2019	Penulis menjelaskan tentang pengertian dari <i>web scraping</i> .
Priyanto & Ma'arif	2018	Penulis menjelaskan konsep perancangan sistem <i>waterfall</i> .
Arsyad, Pramono, Isnawaty, Yamin, & Sarita	2019	Hasil penelitian merupakan sebuah program berbasis website yang dirancang menggunakan konsep <i>waterfall</i>

Penelitian ini mengembangkan sebuah sistem yang berfungsi untuk *scraping website-website* yang berisikan tentang pariwisata seperti yang dilakukan oleh (Darmawiguna et al., 2019), aplikasi yang dirancang menggunakan bahasa pemrograman *C#* seperti penelitian (Asif et al., 2019) serta menggunakan *library Anglesharp* dimana memiliki performa yang baik sesuai dengan hasil evaluasi peneliti (Uzun et al., 2017), dan aplikasi yang dihasilkan ialah berbasis website serta metode yang digunakan dalam pengembangan sistem ialah *waterfall* seperti penelitian (Priyanto & Ma'arif, 2018) dan (Arsyad et al., 2019).

2.2 Landasan Teori

2.2.1 Web Scraping

Menurut (Kadam, Shinde, Sharma, & Mali, 2018) *Web scraping* merupakan sebuah aktifitas mengekstrak kerangka halaman website (*HTML*) dari halaman url yang diakses untuk keperluan mengumpulkan data yang dapat diolah menjadi informasi yang ditujukan untuk keperluan pribadi. Kemudian menurut (Priyanto & Ma'arif, 2018) *web scraping* ialah suatu proses pengambilan seluruh atau sebagian informasi pada sebuah website yang dirancang menggunakan bahasa markup seperti *HTML* atau *XHTML* yang digunakan untuk kepentingan lain.

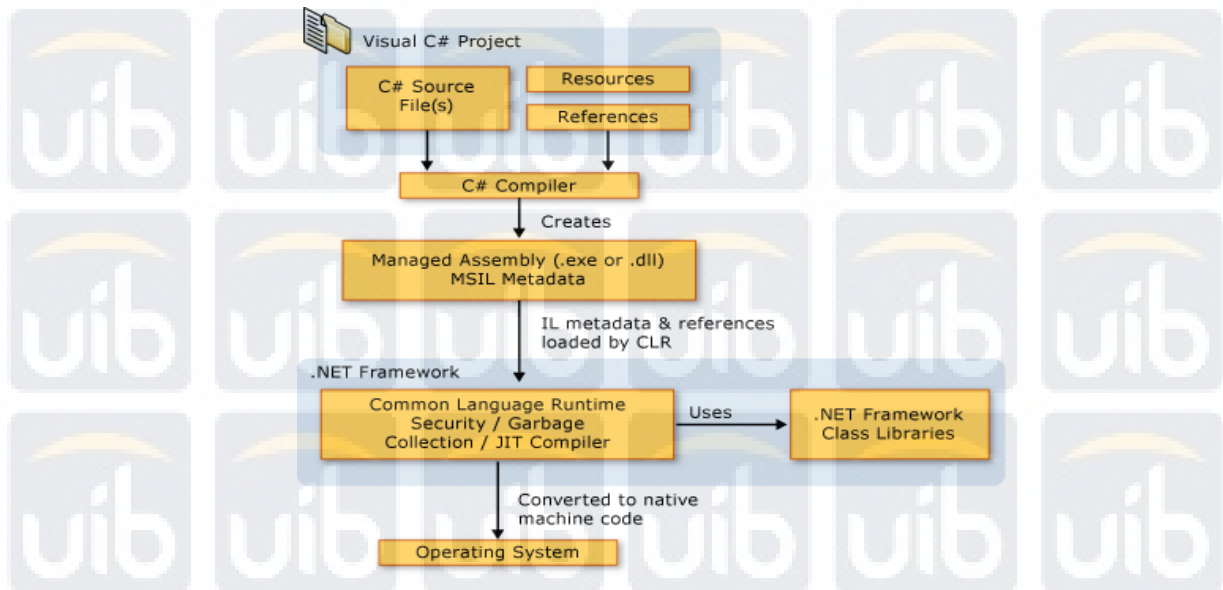
2.2.2 C#

Bahasa pemrograman *C Sharp* atau dikenal dengan *C#* adalah bahasa pemrograman yang dibuat oleh Perusahaan *Microsoft* bersama *.NET* yang tergolong bahasa pemrograman tingkat tinggi yang juga menganut cara kerja berpusat pada objek. Bahasa pemrograman *C#* dapat merancang berbagai macam aplikasi serupa aplikasi berbasis *desktop, games, mobile, serta website* (Al-bastami, 2017).

Bahasa pemrograman *C#* adalah yakni salah satu bahasa pemrograman yang tergolong dalam bahasa tingkat tinggi seperti bahasa pemrograman *Java, Python, Node Js, VB. NET*. Program yang dirancang atau dibangun menggunakan bahasa pemrograman *C #* adalah sistem berorientasi objek. Bahasa pemrograman *C #* dikatakan sebagai bahasa pemrograman berorientasi objek karena ada definisi kelas, metode dan konstruktor yang akan dieksekusi oleh komputer.

Net Framework ialah konstruksi yang merupakan bagian dari *command line* bahasa pemrograman *C#*, kerangka kerja ini berfungsi sebagai wadah yang dimana dipergunakan untuk mengoperasikan kode-kode ataupun instruksi yang dirancang dan menghasilkan proses sesuai dengan rancangan yang diharapkan. Bahasa pemrograman *C#* ini memanfaatkan arsitektur *Common Language Runtime (CLR)*.

Dalam perancangan sistem pada penelitian ini, penulis memanfaatkan IDE (*Integrated Development Environment*) dari *Microsoft* yang sering dinyatakan dengan *Visual Studio 2017 Community* yang berfungsi sebagai perangkat lunak yang menampung kode-kode perintah yang berguna untuk menjalankan sistem yang dirancang. lihat gambar 2.1 tentang pendeskripsian cara kerja IDE *C#*



Gambar 2.1 Cara Kerja IDE

Kelebihan bahasa pemrograman C# sebagai berikut:

1. Bahasa pemrograman tingkat tinggi (mudah dipahami)
2. Berorientasi pada objek (*Object Oriented Programming*)
3. Tersedia *class*, & *library* yang lengkap
4. Fitur *memory auto management*.

2.2.3 ASP.NET Core

ASP.NET CORE adalah suatu kerangka kerja yang dirancang oleh Microsoft yang digunakan sebagai wadah atau tempat untuk membangun suatu program atau aplikasi berbasis website, yang dimana sebelumnya aplikasi berbasis website yang dihasilkan dari bahasa pemrograman C# menggunakan .NET yang menyebabkan aplikasi yang dirancang hanya dapat diimplementasikan atau dihosting pada web server windows yang mendukung IIS (*Internet Information Services*). Dengan adanya *ASP.NET CORE* memungkinkan pengembangan sistem atau aplikasi

website yang menggunakan bahasa pemrograman C# ini dapat diimplementasikan atau dihosting juga pada web server lainnya selain windows, yaitu Linux.

2.2.4 SQL

SQL merupakan kepanjangan dari *Structured Query Language* dimana ini merupakan salah satu bahasa dalam dunia pemrograman yang berguna untuk melakukan manajemen basis data yang menganut konsep RDBMS. Jenis query atau perintah yang terdapat dalam bahasa pemrograman ini dibagi menjadi 3 kelompok yang dipetakan sesuai dengan fungsi atau kegunaannya, yaitu sebagai berikut:

1. *Data Definition Language (DDL)*

Query atau perintah yang digolongkan dalam kelompok ini memiliki fungsi untuk membangun skema sistem basis data dan tabelnya. Berikut contoh perintah *DDL* yaitu, *Create*, *Alter*, *Rename*, dan *Drop*.

2. *Data Manipulation Language (DML)*

Query atau perintah yang digolongkan dalam kelompok ini memiliki fungsi yang berhubungan langsung dengan data yang ada didalam basis data. Contohnya ialah seperti proses untuk memasukkan data baru, menyunting data, menghapus, serta mengambil data yang disimpan dalam basis data. Berikut contoh perintah *DML* yaitu, *Insert*, *Update*, *Delete*, dan *Select*.

3. *Data Control Language (DCL)*

Query atau perintah yang digolongkan dalam kelompok ini memiliki fungsi untuk mengatur tentang hak akses yang dimiliki oleh pengguna dalam pengaksesan basis data. Berikut contoh perintah *DCL* yaitu, *Grant* dan *Revoke*.

Kemudian setelah penjelasan mengenai 3 jenis atau kelompok perintah dalam bahasa pemrograman SQL, terdapat juga 2 jenis tambahan yaitu, *Transactional Control Language (TCL)* jenis perintah ini disebut dengan proses transaksi, dan *query* ini digunakan ketika suatu proses yang memiliki perintah lebih dari 1 didalamnya, dan menerapkan konsep, proses transaksi dianggap berhasil apabila keseluruhan proses didalamnya dieksekusi tanpa ada masalah, jika terdapat masalah maka transaksi tersebut dianggap tidak berhasil, berikut contoh perintah *TCL* yaitu, *Set Transaction*, *Commit*, dan *Rollback*. Kemudian jenis yang keduanya ialah *Programmative SQL* jenis perintah ini berhubungan langsung dengan proses sub program atau dikenal dengan *stored procedure* atau penjelasan mengenai skema basis data, berikut contoh perintahnya yaitu, *Describe*, *Prepare*, *Explain*, dan *Declare*.

2.2.5 HTML

HTML merupakan kepanjangan dari *Hyper Text Markup Language* yang merupakan salah satu bahasa yang ada dalam dunia pemrograman sistem berbasis website. *HTML* ini merupakan bahasa yang disusun atas tanda-tanda tertentu atau umumnya disebut dengan *tag*, contohnya seperti `<head>`, `<body>`, `<div>`, `<table>`, `<form>`, dan sebagainya. *Tag* ini bertujuan untuk menampilkan informasi yang akan dimunculkan dalam laman website.

2.2.6 Javascript

Javascript merupakan salah satu bahasa pemrograman berbasis website juga, yang dimana memiliki fungsi untuk mengelola halaman website perpaduan dari *HTML* dan *CSS* guna membuat halaman website tersebut lebih menarik dan responsif. *Javascript* seiring waktu dikembangkan menjadi lebih beragam atau

beberapa jenis guna membantu sistem developer memilih wadah yang digunakan agar sesuai dengan kebutuhannya. Contohnya *Jquery*, dan *AJAX*.

2.2.7 *Anglesharp*

Anglesharp merupakan salah satu dari *library* yang berguna untuk melakukan proses mengekstrak konten dari suatu laman website. Cara kerja dari *Anglesharp* ini ialah melakukan pengambilan data *parsed DOM* yang berisikan seluruh property atau kerangka website dan tersusun atas node-node yang dimana dapat diakses datanya menggunakan *LINQ* yang membuat ekstraksi menjadi lebih efisien (Uzun et al., 2017).

2.2.8 *Database*

Database atau basis data memiliki pengertian sebagai penampung data yang dimana sebuah tempat penyimpanan seluruh data transaksi kedalam tabel-tabel yang tersusun secara struktur (Gunawan, 2016). *Database* atau basis data disimpan kedalam perangkat keras yang dimana proses penyimpanan atau manipulasi data menggunakan perangkat lunak. Basis data sangat berperan penting dilingkungan sistem informasi, dimana seluruh transaksi atau proses yang terjadi pada suatu sistem itu disimpan ke basis data.

Penyusunan basis data pada umumnya dilakukan dengan cara yang sederhana, yakni data atau transaksi yang terjadi disistem akan dicatat kedalam bentuk tabel agar memudahkan dalam penyusunannya serta lebih mudah dalam penampilan data. Basis data dibagi menjadi 2 macam, sebagai berikut :

1. *Database* hirarki

Basis data hirarki atau dikenal dengan *Database* Hirarki merupakan basis data yang melakukan penyimpanan data menggunakan konsep

seperti hirariki atau pohon. Susunan seperti pohon yang dimana memiliki ranting yang saling berhubungan dan tidak bisa dipisahkan.

2. *Database* relasional

Basis data relasional merupakan basis data yang dimana konsep penyimpanan datanya kedalam bentuk tabel-tabel yang tersusun secara terstruktur. Data pada setiap tabel yang disimpan, dibasis data memiliki hubungan antar tabel. Contoh: dengan kata lain setiap tabel memiliki hubungan atau keterkaitan dengan tabel lainnya yang saling terhubung atau biasa disebut dengan relasi.

Komponen utama dari sistem basis data terdiri atas beberapa bagian, yaitu sebagai berikut:

1. *Database* atau dikenal dengan istilah basis data adalah kumpulan data-data yang tersimpan serta terorganisir, sehingga data menjadi mudah diakses dan dimanipulasi.
2. *Hardware*: sebuah perangkat keras yang dimana berfungsi sebagai media penyimpanan eksternal, dan memiliki piranti output dan input.
3. Sistem operasi ialah suatu sistem yang dirancang dengan tujuan untuk mengoperasikan suatu alat melewati sistem operasi, contohnya sistem operasi pada komputer. Sistem operasi ini memiliki kendali atas sumberdaya pada alat yang dioperasikan menggunakan sistem operasi.
4. *Software*: merupakan suatu perangkat lunak yang digunakan sebagai tujuan untuk menjembatani atau dijadikan sebagai penghubung antara media penyimpanan data perangkat keras dengan sistem computer. Contoh dari perangkat lunak yang digunakan sebagai basis data ialah SQL Server, dan MYSQL.

5. DBMS (*Database Management System*) merupakan suatu perangkat lunak yang dirancang khusus untuk melakukan kegiatan membuat data, mengelola basis data serta manipulasi data. Contoh program

DBMS yang dimaksud ialah SQL Server.

2.2.9 Unified Modeling Language (UML)

Unified Modeling Language (UML) adalah suatu bahasa yang baku atau standart yang pada umumnya UML ini digunakan di dunia industry dengan tujuan agar dapat melakukan pengidentifikasi kebutuhan, melakukan analisa perancangan sistem, dan desain sistem, serta dapat digunakan untuk melakukan kegiatan menggambar arsitektur dalam pemograman berorientasi objek. UML memiliki pengertian suatu bahasa standart yang telah ditetapkan yang berbentuk seperti diagram, dan memiliki tekse pendukung yang digunakan untuk perancangan sistem inforasi(Saiful & Ambarita, 2016).

Perkembangan teknologi pemograman berorientasi objek menyebabkan terjadi suatu standarisasi dalam bahasa pemodelan itu sendiri yang berfungsi sebagai untuk pemodelan visual.

1. Flowchart

Flowchart adalah suatu diagram yang digambar atau dirancang menggunakan kumpulan simbol-simbol yang mempresentasikan bagaimana suatu proses berjalan dengan detail dan menggambarkan juga hubungan antar proses dalam sebuah sistem yang saling berkaitan (Mimin, 2016)

2. Use Case Diagram

Use case diagram merupakan diagram yang merupakan bagian dari UML yang termasuk juga dalam pemodelan visual, kegunaan dari diagram ini ialah untuk melakukan pemodelan kelakuan dari aktor terhadap sistem. *Use*

case ini menggambarkan kejadian atau proses yang terjadi antara aktor dalam sistem. *Use Case* diagram dapat diartikan juga sebagai pemodelan visual atas fungsi yang terdapat dalam sistem, kemudian menggambarkan juga dari fungsi tersebut aktor mana saja yang menggunakannya (Mimin, 2016). *Use case* memiliki beberapa istilah seperti sebagai berikut:

1. Aktor, dikenal sebagai pelaku atau pengguna yang berinteraksi dengan sistem..
2. *Association / Directed Association* adalah hubungan yang statis yang terjadi antar elemen yang dimana suatu elemen ini memiliki atribut yang didapat dari elemen lainnya..
3. *Generalization / Pewarisan* merupakan hubungan yang terjadi antar elemen yang disebut atau dikenal dengan istilah hubungan hierarkis.

3. *Sequence Diagram*

Sequence Diagram adalah suatu diagram yang menjelaskan secara detail hubungan interaksi antara aktor dengan fungsi disistem yang digambarkan pada *use case diagram*. Ketentuan pembuatan *Sequence Diagram* ialah jumlah dari diagram ini harus sebanding dengan *use case diagram* (Mimin, 2016).

4. *Class Diagram*

Diagram kelas atau biasa dikenal dengan istilah *class diagram* merupakan pemodelan visual yang menggambarkan struktur dari sistem yang telah dirancang, dan kelas diagram ini memiliki atribut, metode atau operasi (Mimin, 2016).

5. *Entity Relational Diagram*

ERD merupakan salah satu pemodelan visual yang banyak digunakan untuk merancang pemodelan untuk basis data, ERD dikembangkan berdasarkan teori matematika yang dikenal dengan nama himpunan. ERD ini pemodelan basis data yang bersifat relasional yang dimana setiap entitas dalam pemodelan ini memiliki hubungan (Mimin, 2016).

Komponen penyusun ERD sebagai berikut :

a. Entitas (*entity*)

Entitas dapat diartikan sebagai kumpulan objek yang secara keseluruhan berbeda antar satu objek dengan objek lainnya.

Pemodelan visual ERD, entitas digambarkan sebagai persegi panjang.

b. Relasi (*relational*)

Relasi dalam pemodelan visual ERD adalah suatu hubungan yang terjalin antar satu entitas atau beberapa entitas. Singkatnya relasi dalam pemodelan visual ERD digambarkan menggunakan simbol belah ketupat yang memiliki teks yang secara harfiah berguna untuk menghubungkan 2 entitas atau lebih.

c. Kardinalitas atau Derajat Relasi (*Cardinality / Relationship Degree*)

Kardinalitas merupakan suatu keadaan yang menggambarkan jumlah kemunculan dari sebuah entitas yang dimana berhubungan dengan entitas lainnya. Jenis Derajat Kardinalitas:

i. *One to One* (1:1)

Hubungan 1:1, merupakan kardinalitas atas hubungan yang terjadi antar entitas hanya muncul sekali saja..

ii. *One to Many / Many to One (1: N / N : 1)*

Hubungan 1:N atau N:1, merupakan suatu keadaan kardinalitas dari suatu entitas dengan entitas lainnya yang dapat menghasilkan kemunculan lebih dari satu.

iii. *Many to Many (M : N)*

Hubungan M:N, memiliki pengertian dimana setiap entitas pada basis data dapat memiliki hubungan lebih dari satu atau dapat diartikan setiap entitas dapat berhubungan dengan banyak entitas.

2.2.10 *System Development Life Cycle*

System development life cycle (SDLC) merupakan salah satu metode yang digunakan untuk pengembangan sistem yang dapat dikatakan umum digunakan oleh pengembang sistem. SDLC memiliki beberapa tahap didalamnya saat pengembangan suatu sistem informasi, yakni sebagai berikut, tahap perencanaan sistem, tahap analisa, tahap perancangan dan pengembangan sistem, tahap implementasi sistem dan pemeliharaan pada sistem.

Model *Waterfall* adalah model pengembangan sistem yang merupakan bagian dari SDLC juga, dan pada umumnya *waterfall* dikenal dengan istilah dengan model konvensional. Model ini berjalan dengan sistematis serta urut, dan secara garis besar ada 5 tahapan yaitu sebagai berikut:

1. Analisis Sistem
2. Desain Sistem
3. Pengembangan Sistem

4. Pengujian Sistem

5. Implementasi Sistem

SDLC memiliki tahapan-tahapannya yang terdiri dari sebagai berikut:

1. Tahap menganalisa sistem ialah suatu kegiatan menganalisis sistem yang sudah ada atau sedang digunakan untuk kegiatan operasional perusahaan, dan tujuan dari tahap ini adalah melakukan perancangan sistem baru yang merumuskan dengan cara *fishbones*. *Fishbones* adalah salah metode yang dipraktikan dengan cara pendekatan, dimana mencakup pendekatan pada sumber daya manusia, kerja, mesin, lingkungan, uang dan sistem yang merupakan dari akar permasalahan dalam perancangan sistem (Gunawan, Hariyati, & Fitri, 2018).
2. Tahap mendesain sistem ialah suatu kegiatan mendesain sebuah sistem seperti menentukan bagaimana proses sistem berjalan, aliran data, dan halaman tatap muka sistem berdasarkan hasil dari tahap analisa.
3. Tahap pengembangan sistem ialah suatu kegiatan atau tahap dimana sistem itu sedang dirancang menggunakan kode-kode bahasa pemrograman.
4. Tahap pengujian sistem, ialah suatu kegiatan dimana sistem yang sudah dirancang akan dilakukan uji percobaan, kegiatan ini atau tahap ini dilakukan bertujuan supaya dapat menghindari terjadi kesalahan atau *malfunction* pada sistem yang dirancang saat melakukan implementasi sistem.

5. Tahap implementasi sistem ialah tahap dimana sistem sudah melewati semua proses tahapan untuk perancangan sebuah sistem, pada tahap ini sistem yang dirancang diimplementasikan agar siap dioperasikan.

2.3 Aplikasi Yang Digunakan Dalam Pengembangan Sistem

2.3.1 *Visual Studio 2017 Community*

Microsoft Visual Studio adalah sebuah perangkat lunak yang dikenal sebagai IDE (*Integrated Development Environment*) yang digunakan sebagai wadah untuk pengembangan sistem. Perangkat lunak ini dapat mengembangkan atau merancang berbagai macam aplikasi yang dijalankan dalam berbagai platform, yakni seperti aplikasi mobile atau android, website, dan aplikasi desktop..

Teknologi *.NET Framework* merupakan kerangka kerja yang dirancang oleh perusahaan *Microsoft* sebagai wadah untuk developer agar dapat merancang sistem menggunakan perangkat lunak tersebut. Kerangka kerja ini mendukung *native code* atau biasa dikenal dengan bahasa mesin.

2.3.2 *Microsoft SQL Server*

Microsoft SQL Server termasuk dalam *Relational Database Management System* (RDBMS) merupakan produk dari *Microsoft*. Bahasa yang digunakan dalam program *SQL Server* ini adalah *Transact-SQL* yang merupakan adaptasi dari SQL Standar ANSI/ISO. Perangkat lunak *SQL server* digunakan untuk manipulasi basis data ini umumnya digunakan oleh perusahaan yang berskala kecil hingga kelas menengah (Irawan, Misdram, & Aini, 2017).

SQL Server merupakan perangkat lunak yang dirancang untuk mengelola basis data atau biasanya dikenal umum oleh masyarakat luas dengan nama RDBMS (*Relational Database Management System*).

Kelebihan *SQL Server* sebagai berikut:

1. Pengelolaan kata sandi yang baik.
2. Memiliki fitur pemulihan basis data dan backup data.
3. Membuat clustering pada kolom data.
4. Dapat diimplementasikan pada skala menengah kebawah hingga atas

