

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Objek Penelitian

Penelitian ini menggunakan objek tentang data pemain *NBA* pada tahun 2022-2023. Data statistik pemain *NBA* yang digunakan dalam penelitian ini diperoleh dari *website* resmi *NBA* ([www.NBA.com](http://www.NBA.com)). *NBA* sendiri menyediakan seluruh data statistik pemain *NBA* yang terdaftar secara resmi dan bermain pada saat *season* tersebut. *Season* disini diselenggarakan setiap tahunnya dan memiliki beberapa tipe pertandingan yang dimulai dari *Preseason Game*, *Regular Season Game*, *Playoffs Game*, dan *All-Star Game*. Pada penelitian ini data statistik pemain yang akan digunakan adalah data statistik pemain pada saat *Regular Season* yaitu berisikan 25 fitur antara lain *minutes\_played*, *field\_goals\_made*, *free\_throws\_attempted*, *double\_doubles*, *assists*, *defensive\_rebounds*, *points*, *rebounds*, *field\_goals\_attempted*, *free\_throws\_made*, *steals*, *turnovers*, *blocks*, *3\_point\_attempted*, *field\_goal\_percentage*, *3\_point\_percentage*, *games\_played*, *wins*, *personal\_fouls*, *offensive\_rebounds*, *loses*, *3\_point\_made*, *age*, *free\_throw\_percentage*, *triple\_doubles*. Seluruh data statistik di atas akan menjadi dasar perbandingan bagi para pemain untuk mendapatkan gelar *Most valuable player (MVP)* pada *season* tersebut.

Pemilihan sumber untuk data yang digunakan pada penelitian ini merupakan sumber yang valid dan terpercaya. *Website* resmi *NBA* yaitu *NBA.com* menyediakan berbagai informasi lengkap seputar liga *NBA* mulai dari Games, Schedule, Watch, News, Stats, Standings, Teams, dan Players. Informasi yang diberikan didalam *website* resmi *NBA* disini sudah dipastikan valid dan dapat dipercaya kebenarannya. Oleh karena itu, Penelitian ini akan menggunakan data statistik (*stats*) para pemain pada tahun 2022-2023 yang diperoleh dari *website NBA.com* untuk digunakan sebagai data utama dalam penelitian mengenai prediksi gelar *MVP* yang akan diberikan pada salah satu pemain *NBA* dengan statistik terbaik.

### 3.1.1 Profil Perusahaan

*National Basketball Association* atau biasa disebut *NBA* merupakan liga bola basket professional yang bertempat di Amerika Serikat. *NBA* sendiri merupakan salah satu liga bola basket terbesar dan paling terkenal diseluruh dunia. Liga ini diikuti oleh 30 tim yang mana 29 tim berasal dari Amerika Serikat dan 1 tim berasal dari Kanada. Liga bola basket *NBA* dibentuk dan didirikan pada tanggal 6 Juni 1946 di *New York*. Pertama kali didirikan *NBA* memiliki nama awal yaitu *Basketball Association of America (BAA)* dan diubah menjadi *NBA* pada tanggal 3 Agustus 1949. [26]

Liga *NBA* dilaksanakan setiap musim. Pada umumnya musim regular *NBA* akan diselenggarakan dari bulan Oktober hingga April dengan jumlah total pertandingan sebanyak 82 pertandingan setiap masing masing tim yang bermain di liga *NBA*. Setelah musim regular selesai diselenggarakan maka akan diadakan babak playoffs. Babak Playoffs disini hanya akan mengambil 8 team dengan peringkat tertinggi dari masing masing wilayahnya (*West & East*) dengan ini babak playoffs akan diikuti oleh 16 tim saja. 16 tim ini akan bertanding untuk memperebutkan gelar juara yaitu *Larry O'Brien Championship Trophy*. [27]

Liga *NBA* dibagi menjadi dua wilayah yaitu wilayah Timur (*East*) dan wilayah Barat (*West*). Berikut merupakan daftar tim *NBA* pada setiap wilayah:

#### **Wilayah Timur (*East*)**

- *Boston Celtics*
- *Brooklyn Nets*
- *New York Knicks*
- *Philadelphia 76ers*
- *Toronto Raptors*
- *Chicago Bulls*
- *Cleveland Cavaliers*

- *Detroit Pistons*
- *Indiana Pacers*
- *Milwaukee Bucks*
- *Atlanta Hawks*
- *Charlotte Hornets*
- *Miami Heat*
- *Orlando Magic*
- *Washington Wizards*

#### **Wilayah Barat (West)**

- *Denver Nuggets*
- *Minnesota Timberwolves*
- *Oklahoma City Thunder*
- *Portland Trail Blazers*
- *Utah Jazz*
- *Golden State Warriors*
- *Los Angeles Clippers*
- *Los Angeles Lakers*
- *Phoenix Suns*
- *Sacramento Kings*
- *Dallas Mavericks*
- *Houston Rockets*
- *Memphis Grizzlies*
- *New Orleans Pelicans*
- *San Antonio Spurs*

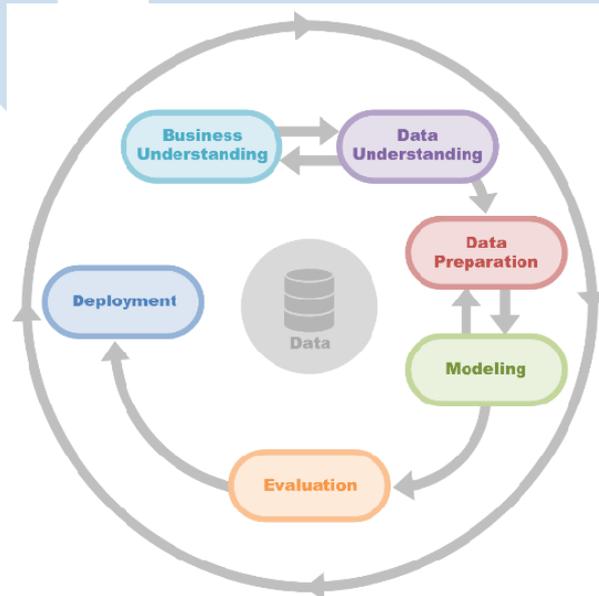
### **3.2 Metode Penelitian**

Pada penelitian ini akan digunakan 3 Algoritma yang berbeda untuk mendapatkan 3 hasil prediksi untuk dibandingkan satu sama lain. 3 Algoritma yang akan digunakan adalah *Decision Tree*, *Linear Regression*, dan *Support*

*Vector Regression*. Selain itu adapun *Framework* yang digunakan pada penelitian ini yaitu *Framework CRISP-DM*.

### 3.2.1 CRISP-DM

The Cross Industry Standard Process for *Data Mining* atau biasa disebut *CRISP-DM* merupakan sebuah model proses *data mining / datamining Framework* yang berguna sebagai dasar untuk melakukan proses *data science* atau *data mining*. Didalam *CRISP-DM* terdapat 6 tahapan yang harus dilakukan secara berurutan yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. Keenam tahapan disini harus dipahami dan dilakukan secara berurutan agar dapat menghasilkan hasil yang terbaik dan akurat pada proyek *data mining*. [15]



Gambar 3. 1 CRISP - DM

#### 3.2.1.1 Business Understanding

*Business Understanding* adalah tahapan pertama dalam *CRISP-DM*. Dalam tahap kali ini akan dilakukan proses pemahaman suatu bisnis serta penggunaan data mining yang akan digunakan didalamnya. Pada penelitian ini, data mining digunakan untuk membantu media dalam melakukan prediksi

terhadap pemain yang mendapatkan gelar *MVP* pada musim tertentu pada liga *NBA*.

### **3.2.1.2 Data Understanding**

*Data Understanding* merupakan tahap pemahaman data. Dalam tahap ini data akan dikumpulkan dan diperiksa kualitasnya. Dalam penelitian ini data statistik pemain *NBA* akan diperoleh dari *website* resmi *NBA* yaitu *NBA.com*. Setelah itu pengecekan data akan dilakukan mulai dari *variable* dan *field* yang ada didalam data statistik. Data statistik yang diperoleh dan digunakan dalam penelitian ini tentunya valid dan terpercaya karena dikumpulkan dari sumber resmi *NBA* itu sendiri.

### **3.2.1.3 Data Preparation**

*Data Preparation* merupakan tahapan dimana data akan disiapkan menjadi data yang dapat digunakan secara matang dan akurat. Dalam penelitian ini, data statistik *NBA* akan disiapkan secara terperinci mulai dari proses pembersihan data dari berbagai faktor yang dapat merusak tingkat kestabilan data, filterisasi pada data dimana *field* atau *variable* yang tidak berguna dapat disingkirkan. Tahapan ini sangatlah dibutuhkan dalam proses analisa karena dengan ini maka data dapat dipastikan kualitasnya dan tentunya sudah siap untuk digunakan dalam sebuah penelitian. Tidak hanya itu, kesiapan data juga akan mempengaruhi tingkat akurasi yang akan dihasilkan nantinya.

### **3.2.1.4 Modeling**

*Modeling* merupakan tahapan pemilihan algoritma dan pembentukan model dalam sebuah penelitian. Pada tahap ini model *Machine Learning* akan dibuat dengan algoritma yang sudah ditentukan. Dalam penelitian ini, model *Machine Learning* akan dibuat dan dibentuk menggunakan 3 algoritma yang berbeda yaitu *Decision Tree*, *Linear Regression*, dan *Support Vector Regression*.

### 3.2.1.5 Evaluation

*Evaluation* merupakan tahapan evaluasi terhadap model yang sudah dibentuk pada tahap *Modeling*. Pada tahap ini ketiga model akan dievaluasi secara terperinci. Pada penelitian ini, ketiga model akan diuji dan dibandingkan satu sama lain. Hal ini dilakukan untuk dapat melihat apakah model algoritma yang dibentuk sudah sesuai atau belum dan juga melihat model dengan algoritma mana yang paling sesuai dan memiliki hasil akurasi terbaik.

### 3.2.1.6 Deployment

*Deployment* merupakan tahapan akhir dari *CRISP-DM*. Pada tahap ini model yang paling sesuai atau yang terbaik akan diterapkan pada data. Tidak hanya itu, tahap ini juga mencakup proses pemantauan pada model yang digunakan. Hal ini dilakukan agar dapat mengurangi resiko terjadinya masalah dan juga mengantisipasi masalah yang akan timbul saat proses penerapan berlangsung. Pada penelitian ini, model yang sudah dibuat dan dievaluasi akan diterapkan pada data statistik *NBA* serta akan dilakukan pemantauan secara berlanjut untuk mengawasi proses agar dapat menghasilkan hasil akurasi yang tinggi dan akurat.

## 3.3 Teknik Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang didapatkan dari *website* resmi *NBA* yaitu *NBA.com*. Data yang peroleh dan digunakan pada penelitian ini merupakan data statistik pemain yang aktif dan bermain dalam liga *NBA* pada tahun 2022-2023 dengan tipe regular *season* games. Total data yang diperoleh dan digunakan pada penelitian ini adalah sebanyak 468 data pemain.

## 3.4 Teknik Pengambilan Sampel

Populasi dari data yang diambil adalah keseluruhan pemain yang terdaftar secara resmi dan bermain dalam liga *NBA* pada tahun 2022-2023 yaitu sebanyak 505 pemain. Untuk sample yang digunakan pada penelitian ini adalah 468 dari

populasi. Sample yang digunakan adalah pemain yang sedang aktif bermain di liga *NBA* pada musim 2022 – 2023.

### 3.5 Teknik Analisis Data

Analisis data yang dilakukan didalam penelitian ini dimulai dari proses pengambilan data pemain *NBA* yang aktif bermain dan secara resmi terdaftar sebagai pemain dalam liga *NBA*. Data statistik pemain *NBA* disini diperoleh melalui *website* resmis *NBA* yaitu *NBA.com* yang mana menyediakan seluruh data statistik para pemain di setiap tahunnya. Data yang diperoleh dan digunakan dalam penelitian ini merupakan data statistik pemain *NBA* yang aktif bermain pada tahun 2022-2023 dalam pertandingan regular di musim tersebut. Data Statistik pemain disini memiliki total records sebanyak 468 yang berarti didalam data statistik yang diperoleh terdapat 468 pemain aktif yang bermain didalam liga *NBA* pada tahun 2022-2023. Data statistik pemain disini terdiri dari beberapa fields yaitu *Player, Team, Games Played, Minutes Played, Points, Field Goals Made, Field Goals Attempted, Field Goal Percentage, 3 Point Field Goals Made, 3 Point Field Goals Attempted, 3 Point Field Goal Percentage, Free Throws Made, Free Throws Attempted, Free Throw Percentage, Offensive Rebounds, Defensive Rebounds, Defensive Rebounds, Assists, Steals, Blocks, Turnovers*. Selanjutnya data akan diproses dan dikelola secara bertahap mulai dari persiapan data, pembersihan data, hingga pengecekan format data. Hal ini dilakukan agar memastikan bahwa data memiliki kualitas yang baik serta siap untuk digunakan didalam proses penelitian. Setelah data siap digunakan maka data akan dibagi menjadi dua tipe data yaitu data testing dan data training. Tahapan selanjutnya adalah proses pemodelan data menggunakan algoritma *Machine Learning* dengan teknik regresi. Algoritma yang digunakan didalam penelitian ini adalah *Decision Tree, Linear Regression, dan Support Vector Regression*. Proses tahapan selanjutnya adalah melakukan regresi pada data statistik pemain dimana nantinya akan ada perbandingan pengaruh dari seluruh variable yang ada didalam data pada tingkat kesempatan pemain untuk mendapatkan gelar *MVP*. Tahapan ini akan dilakukan dengan menggunakan 3 algoritma berbeda yang mana setiap algoritma akan menghasilkan tingkat akurasi

dalam melakukan regresi pada para pemain yang memiliki kesempatan untuk mendapatkan gelar *MVP* pada musim 2022-2023 berdasarkan statistiknya.

### 3.6 Bahasa Pemrograman, Tools, dan Framework

#### 3.6.1 Bahasa Pemrograman

Penelitian ini memilih *Python* sebagai bahasa pemrograman yang akan digunakan dalam proses analisa prediksi dalam penelitian ini. *Python* dipilih karena fitur dan fleksibilitas yang dimilikinya serta penggunaan *Python* tergolong cukup mudah untuk dipahami. Akses untuk menggunakan dan menerapkan bahasa pemrograman *Python* juga sangat mudah. Hal ini dikarenakan *Python* berbasis *open-source* atau dengan kata lain dapat digunakan dengan gratis tanpa harus mengeluarkan biaya apapun. Selain itu, *Python* juga menyediakan *Library* yang cukup lengkap dan juga sintaks yang ada pada bahasa pemrograman *Python* tergolong sederhana. Hal ini membuat bahasa pemrograman *Python* digemari oleh banyak analis untuk menopang mereka dalam melakukan penelitian analisa pada sebuah data dengan ukuran yang besar. [24]

Penggunaan bahasa pemrograman *Python* pada penelitian ini didasarkan oleh fitur dan juga kegunaan bahasa pemrograman *Python* yang dapat dikatakan lebih unggul daripada bahasa pemrograman lainnya seperti R dan Matlab. Berikut merupakan *table* perbandingan *Python*, R dan Matlab.

**Tabel 3. 1 Python, R, and Matlab Comparison Table**

<i>Python</i>	<b>R</b>	<b>Matlab</b>
<i>Open-source</i>	<i>Open-source</i>	Not Open Source
<i>User Friendly</i>	<i>User Friendly</i>	More Complicated
Easy to Access	Easy to Access	Less Easy to Access
Free	Free	Not 100% Free
Fast	Less Fast	Normal
Simple Syntax	Good <i>Visualization</i>	Good Graph for Complex Mathematics

Perbandingan tabel di atas memperlihatkan keunggulan dan kekurangan dari ketiga bahasa pemrograman yang dapat digunakan pada penelitian ini.

Tetapi pada akhirnya, penelitian ini memutuskan untuk menggunakan bahasa pemrograman *Python* karena dinilai paling sesuai dan tepat untuk penelitian ini. Bahasa pemrograman *Python* memiliki berbagai keunggulan dan kegunaan yang dibutuhkan. Salah satu yang terpenting adalah *Python* berbasis *open-source* dan mudah untuk digunakan. Selain itu, *Python* juga tergolong cepat dan memiliki syntax yang sederhana dan mudah untuk dimengerti oleh manusia.

### 3.6.2 Tools Deployment

Penelitian ini akan memberikan hasil luaran berbentuk visualisasi *dashboard* dan reports. Hasil visualisasi akan dibuat dengan menggunakan *python*. Hasil visualisasi nantinya akan ditampilkan kedalam sebuah aplikasi berbentuk website. Pada tahap *deployment* ini akan menggunakan *Streamlit*. *Streamlit* merupakan sebuah *framework open-source* yang berguna untuk membangun sebuah web interface terutama untuk aplikasi machine learning atau data science. Penggunaan *streamlit* ini dipilih karena dapat digunakan secara mudah dan cepat. Dengan menggunakan *streamlit*, maka user dapat membangun sebuah web dengan mudah menggunakan bahasa pemrograman *python*. Berikut kelebihan dan kekurangan dari *Streamlit*.

Tabel 3. 2 *Streamlit* Advantages and Disadvantages

Kelebihan	Kekurangan
Sederhana dan mudah	Tidak flexible dengan kostumisasi
Ringan dan cepat	Skalabilitas yang kecil
Interaktif yang dinamis	Dokumentasi yang tergolong sedikit
Integrasi mudah dengan <i>Python</i>	Tidak disarankan untuk aplikasi yang memiliki tingkat kompleksitas tinggi
Deployment mudah dan gratis	Fitur yang minim

Berdasarkan Tabel 3.2 di atas dapat dikatakan bahwa *streamlit* merupakan salah satu *framework* yang tergolong bagus untuk digunakan dalam proses pengembangan aplikasi berbasis *web* sederhana. Hal ini dikarenakan oleh penggunaan *streamlit* yang tergolong mudah, ringan, dan cepat sehingga peneliti dengan basis bahasa pemrograman *python* dapat tetap menggunakan

*streamlit* dan membangun sebuah *web interface* untuk menampilkan hasil visualisasi yang dimilikinya.

### 3.6.3 Framework

Dalam penelitian ini, pemilihan kerangka kerja yang tepat menjadi hal krusial untuk memastikan hasil yang akurat dan relevan. Oleh karena itu, peneliti memutuskan untuk menggunakan *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) sebagai framework utama dalam melakukan prediksi *Most Valuable Player (MVP)* pada pemain *NBA*. Tabel 3.3 merupakan tabel perbandingan framework yang seringkali digunakan yaitu *CRISP-DM*, *KDD*, dan *SEMMA*.

Tabel 3. 3 Perbandingan Framework

CRISP-DM	KDD	SEMMA
Sangat iteratif dengan umpan balik yang sering	Kurang iteratif, lebih linier	Iteratif
Lebih lincah, mudah beradaptasi dengan perubahan	Kurang fleksibel, lebih struktural	Fleksibilitas moderat
Fokus seimbang pada bisnis dan data	Fokus berat pada praproses data	Fokus kuat pada manipulasi data
Proses terstruktur namun fleksibel	Langkah praproses data yang detail	Penekanan kuat pada analisis data eksploratif
Fase yang jelas untuk pemahaman bisnis dan data	Fokus pada penemuan pola	Penekanan pada modifikasi data

*CRISP-DM* dipilih sebagai kerangka kerja untuk penelitian ini karena kerangka ini sangat iteratif dengan umpan balik yang sering, yang memungkinkan penyesuaian dan perbaikan terus-menerus selama proses penelitian. Fleksibilitas *CRISP-DM* yang tinggi membuatnya mudah beradaptasi dengan perubahan, menjadikannya lebih lincah dibandingkan dengan *KDD* yang lebih linier dan kurang fleksibel serta *SEMMA* yang memiliki fleksibilitas moderat. Fokus seimbang *CRISP-DM* pada bisnis dan data memastikan bahwa model prediksi yang dihasilkan tidak hanya akurat dari segi statistik, tetapi juga relevan secara bisnis dan praktis digunakan. Proses yang terstruktur namun fleksibel serta fase-fase yang jelas untuk

pemahaman bisnis dan data dalam *CRISP-DM* membantu dalam mengidentifikasi dan memahami secara mendalam berbagai faktor yang mempengaruhi pemilihan *Most Valuable Player (MVP)* di *NBA*. Pendekatan ini memungkinkan peneliti untuk secara sistematis mengembangkan, menguji, dan memvalidasi model prediksi dengan mempertimbangkan aspek bisnis dan teknis secara komprehensif.

