



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

METODOLOGI PENELITIAN

Penelitian ini menggunakan data pemain dari website resmi NBA pada tahun 2016-2018. Data yang diperoleh kemudian dipindahkan ke Excel untuk diolah kembali sehingga bisa digunakan untuk melakukan prediksi dan visualisasi.

3.1. Objek Penelitian

Penelitian ini dilakukan dengan menggunakan data yang didapatkan dari Portal Resmi data pemain di liga NBA (www.nba.com). Data yang akan diambil berupa data setiap pemain yang terdaftar dan bermain pada liga NBA pada tahun 2016 -2018 seperti *Name, games played, minutes, points, field goal made, field goal attempt, field goal percentage, 3 point made, 3 point attempt, 3 point percentage, free throw made, free throw attempt, free throw percentage, offensive rebound, defensive rebound, rebound, assist, steal, block, turnover, personal foul, efficient*.

Website NBA.com merupakan website resmi yang berisikan informasi mengenai data pemain dan tim yang ada di liga tersebut. Website NBA.com dipilih karena website tersebut merupakan website resmi yang dimiliki oleh liga nba, sehingga dipastikan data yang dimiliki pada website tersebut akurat dan valid.

Pada penelitian ini, hasil prediksi yang diperoleh akan di tampilkan pada *dashboard* visualisasi sehingga dapat dilihat orang banyak dengan mudah dan

dimengerti dengan cepat. Hasil prediksi berupa perbandingan antara 2 tahun MVP dimana perbandingan itu akan membandingkan 10 pemain.

3.2. Pengubahan Format Data

Data pemain yang didapat dari situs NBA berupa excel, dimana data tersebut langsung diambil dari website dan di pindahkan ke excel. Perubahan pada format data tidak perlu dilakukan karena data sudah cocok untuk diolah, namun harus dilakukan perapian data karena ketika dipindahkan dari situs ke dalam Excel, banyak data yang berantakan sehingga harus dilakukan perapian untuk lanjut ke tahap selanjutnya.

3.3. Metode Prediksi

Tabel 3.2. merupakan perbandingan antara metode Linear Regression dan Decision Tree

Tabel 3.1. Tabel Perbandingan Algoritma

<i>Algorithm</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
<i>Linear Regression</i>	<i>Regression</i>	<i>Yes</i>	<i>Lower</i>
<i>Decision Tree</i>	<i>Either</i>	<i>Somewhat</i>	<i>Lower</i>

Penelitian ini menggunakan metode Multiple Linear Regression karena metode ini dianggap yang paling cocok untuk digunakan didalam penelitian ini. Metode ini menggunakan variable sebab akibat dimana sesuai dengan variable yang dimiliki oleh penelitian ini. Metode ini menjelaskan hubungan antara variabel

respon (variabel dependen) dengan faktor-faktor yang mempengaruhi lebih dari satu prediktor (variabel independen).

Regresi linier berganda hampir sama dengan regresi linier sederhana, hanya saja pada regresi linier berganda variabel bebasnya lebih dari satu variabel penduga. Tujuan analisis regresi linier berganda adalah untuk mengukur intensitas hubungan antara dua variabel atau lebih dan membuat prediksi perkiraan nilai y atas x .

Secara umum model regresi linier berganda adalah sebagai berikut :

$$Y = a + b_1X_1 + b_2X_2 + ..$$

Rumus 3. 1. Rumus Multiple Linear Regression

Dimana:

Y = *Variabel Response*)

X = *Variabel Predictor*

a = konstanta

b = koefisien regresi (kemiringan)

Nilai-nilai a dan b dapat dihitung rumus :

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

Rumus 3. 2. Linear Regression a

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Rumus 3. 3. Linear Regression b

Dalam penelitian ini *variable X* atau *predictor* yang akan digunakan adalah *Points, field goal made, field goal attempt, free throw made, free throw attempt, block, steal, rebound, assist, turn over*. Sementara *variable y* atau *dependent* yang digunakan adalah *efficient*.

Penelitian ini memilih *variable efficiency* sebagai *variable dependent* dikarenakan *variable* tersebut merupakan ukuran keberhasilan seseorang menjadi seorang MVP dalam keseluruhan data, dimana *efficiency* mencakup beberapa data dari setiap pemain. Untuk mendapatkan *efficiency* rumus yang digunakan adalah: $((\text{Points} + \text{Rebounds} + \text{Assists} + \text{Steals} + \text{Blocks}) - ((\text{Field Goals Att.} - \text{Field Goals Made}) + \text{Turnovers}))$.

Pada penelitian ini menggunakan 2 tambahan *variable* dari 8 variabel yang digunakan untuk menghitung *eff*. Variabel yang ditambahkan yakni *free throw attempt* dan *free throw made*. Alasan memasukan 2 variabel tambahan adalah untuk membuktikan bahwa kedua atribut tersebut juga memiliki dampak yang besar terhadap penelitian ini.

3.4. Metode Visualisasi

Tabel 3.2. Tabel Perbandingan Metode Visualisasi

Metode	<i>Perbandingan</i>	<i>Proses</i>
<i>Visual Data mining</i>	Pada VDM memiliki 8 fase didalam proses pengerjaanya sehingga proses pengerjaan lebih detil dan jelas	8
<i>User Centrered Design</i>	User Centered Design menggunakan 5 proses didalam pengerjaannya serta USD mementingkan user sebagai pusat dari design	5

Visualisasi yang dilakukan pada penelitian ini bertujuan agar memudahkan bagi para penggemar bola basket untuk melihat data keseluruhan setiap pemain dengan grafik, dimana data-data tersebut dapat dibandingkan dengan antar pemain.

Metode visualisasi yang digunakan untuk melakukan penelitian ini adalah Visual Data Mining (VDM). Berikut langkah-langkah VDM didalam penelitian :

A. *Justify and Planning the Data Visualization and Data minig Project*

Langkah pertama dalam visual *data minig* adalah menentukan lama waktu pengerjaan, tipe proyek, dan juga menentukan roles pada setiap anggota tim. Pada tahap ini akan dilakukan penentuan tipe proyek yang akan

diterapkan pada penelitian ini. Dimana penelitian ini akan menggunakan *production visual data minig*

B. *Identifying the Top Business Questions:*

Menentukan pertanyaan bisnis sehingga arah project menjadi jelas, tidak melebar dan terarah pada tujuan ini. Pertanyaan bisnis pada penelitian ini adalah bagaimana

C. *Choosing the Data:*

Memilih *dataset* yang ingin digunakan yang berhubungan dengan permasalahan bisnis, *dataset* yang akan digunakan merupakan data setiap pemain NBA yang akan bermain di liga tersebut pada tahun 2017- 2018.

D. *Transforming Data Set*

Menginvestigasi data dan melakukan *Transformasi* data dengan menggunakan *logical Transformation* pada dataset.

E. *Verifying the Data Set*

Memeriksa dan memastikan kembali data dan teraplikasikan dengan benar, dan tidak memiliki *error* lagi.

F. *Choosing the Visualization or Data minig Tools*

Memilih *Tools* untuk visualisasi atau *minig Tools* yang dapat menjawab pertanyaan bisnis. Pada penelitian ini akan menggunakan *Microsoft PowerBI* sebagai *Tools*.

G. *Analyzing the Visualization or Data minig Model*

Menganalisa dan menginvestigasi apakah model dari visualisasi dan mining sudah dapat menjawab pertanyaan bisnis atau belum, sehingga dapat menentukan model terbaik yang dapat digunakan untuk menjawab pertanyaan bisnis.

H. *Verify and Presenting Analysis*

Menampilkan visualisasi data ataupun *data mining* untuk pembuatan keputusan.

3.5. **Pemilihan Tools**

Berikut merupakan *Tools* prediksi yang akan digunakan untuk melakukan prediksi dan visualisasi

3.5.1. **Tools Prediksi**

Pada penelitian ini, tools prediksi yang akan digunakan adalah IBM SPSS karena IBM SPSS memiliki analisis metode multiple linear regression serta user interface yang friendly sehingga penggunaan IBM SPSS tidak rumit dan cepat dimengerti. IBM SPSS merupakan sebuah program *computer statistic* yang berfungsi untuk membantu memproses data- data statistik secara tepat dan cepat. SPSS telah digunakan di berbagai bidang persoalan seperti riset pasar, pengendalian dan perbaikan mutu, serta riset-riset sains. Berdasarkan sudut pandang statistik, terdapat dua jenis data yang dapat diolah menggunakan program SPSS, yaitu data kualitatif dan data kuantitatif.

Tabel 3. 3. Perbandingan Tools Prediksi

Tools/Software	RapidMiner	SPSS
Small-Medium Business	√	√
Large Enterprises	√	√
Platform		
On Premise	√	√
Online		
Credit Scoring		
Fraud Detection		
Quality Assurance		
Supply Chain Management		√
Customer Analytic	√	√
Maintenance	√	√
Sales Analytics	√	√

Data kualitatif merupakan data yang dapat dinyatakan dalam bentuk bukan angka, misalnya jenis pekerjaan seseorang yang meliputi nelayan, petani, pegawai, dan lain-lain. Selain itu bisa juga data gender (pria atau wanita), tingkat kepuasan seseorang mulai dari tidak puas, cukup puas, dan sangat puas, dan data lain yang berbentuk bukan angka.

Data kualitatif seperti ini harus dikuantifikasi terlebih dahulu agar dapat diolah dengan statistik. Cara mengkuantifikasikan data kualitatif ini yaitu dengan cara memberi skor tertentu (wanita diberi skor 1, pria diberi skor 2), memberi rangking (tidak puas 1, cukup puas 2, dan sebagainya), atau memberi pendapat (Ya 1, Tidak 2). Ini akan memudahkan data untuk diolah dengan program SPSS.

3.5.2. Tools Visualisasi

Tabel 3. 4. Perbandingan Tools Visualisasi

<i>Tools/Software</i>	<i>Tableau</i>	<i>PowerBI</i>	<i>Qlik</i>
<i>Features</i>			
<i>Business Intelligence</i>	√	√	√
<i>Data Visualization</i>	√	√	√
<i>Dashboard Creation</i>	√	√	√
<i>Customize reporting</i>	√		√
<i>Rating</i>	93%	99%	98%

Pada penelitian ini tools visualisasi yang digunakan adalah *Microsoft Power BI*. *Microsoft Power BI* merupakan aplikasi yang free, sehingga sangat mudah untuk didapatkan. *Microsoft power BI* juga menyediakan fitur- fitur tambahan seperti dashboard creation. *Microsoft power BI* juga memiliki keunggulan didalam customize dashboard sehingga memudahkan dalam hal analisis data yang banyak.

Power BI juga sudah didukung web based sehingga memudahkan pengguna untuk mengunggah hasil *dashboard* ke dalam web agar dapat di share ke pengguna lainnya sehingga visualisasi dapat dibagikan kepada pengguna internet.

3.6. Penelitian Terdahulu

Tabel 3.5. Penelitian Terdahulu

No	Judul Penelitian	Penulis	Tahun	Deskripsi
1	Metode Regresi Linier untuk Prediksi KebutuhanEnergi Listrik Jangka Panjang (Studi Kasus Provinsi Lampung)	M. Syafruddin, Lukmanul Hakim, Dikpride Despa,	2014	Penelitian ini dilakukan dikarenakan meningkatnya pembangunan di Provinsi Lampung, terutama di sector perumahan baik sederhana maupun rumah mewah yang membawa konsekuensi logis berupa peningkatan kebutuhan tenaga listrik. Penelitian ini bertujuan untuk memprediksi kebutuhan energi listrik di Provinsi Lampung hingga Tahun 2030, diharapkan dapatdijadikan sebagai masukkan dalam melakukan perencanaan pembangunan sistemtenaga listrik. Hasil penelitian ini adalah Konsumsi energi listrik di Provinsi Lampung setiap tahunnya mengalami peningkatan yang tidak terlalu signifikan dibandingkan tahun 2013, pada tahun 2014 hasil prediksi konsumsi energi listrik di Provinsi Lampung sebesar 3.986,87 Gwh, pada tahun 2023 prediksi konsumsi energi listrik meningkat mejadi 5.934,98 Gwh, atau menigkat sebesar 26,50% (rata-rata peningkatan konsumsi energi listrik di Provinsi Lampung sebesar 3,83%)
2	Perbandingan 3 Metode Dalam Data Mining Untuk Menentukan Strategi Penjualan Produk Makanan dan Minuman Pada Toserba Lestari Baru	Deny Arga Felani.	2005	Penelitian ini bertujuan untuk membandingkan 3 metode dalam data mining yaitu metode decision tree, k-means, dan linear regression. Penggunaan metode decision tree, K-means (clustering) dan regresi lineardiimplementasikan menggunakan aplikasi RapidMiner, yang nantinya akan dilakukan analisis dari masing-masing metode tersebut untuk menentukan strategi

No	Judul Penelitian	Penulis	Tahun	Deskripsi
				penjualan pada Toserba Lestari Baru Gemolong. Penelitian ini dilakukan dengan sekelompok data untuk mengetahui persentase nilai precision, recall dan accuracy. Hasil dari penelitian tersebut bahwa metode K-Means (clustering) memiliki nilai lebih baik dari metode yang lain pada sisi precision dan accuracy, sedangkan metode Regresi Linier memiliki nilai recall lebih baik dari metode yang lain.

Berdasarkan Tabel 3. 6. penelitian yang telah diadopsi adalah bagaimana cara melakukan metode prediksi dengan menggunakan *multiple linear regression*, dengan mengadopsi persamaan dari *multiple linear regression* serta cara pengerjaannya. Pada penelitian (Felani, 2005) mengadopsi perbandingan antara metode prediksi sehingga penelitian ini akhirnya menggunakan metode *multiple linear regression*.