

II. TINJAUAN PUSTAKA

Banyak penelitian-penelitian yang sebelumnya dilakukan oleh para peneliti terkait tema NBA. Bab ini akan menjabarkan semua penelitian-penelitian terkait yang menjadi rujukan untuk penelitian ini. Seperti yang dilakukan Lee et al [12], playoff merupakan pertandingan NBA setelah musim reguler yang mempertemukan tim peringkat atas bagian barat dan timur untuk memperebutkan juara NBA. Paper ini menggunakan Surprisingly Popular Algorithm untuk menentukan pemenang dari pertandingan basket NBA, khususnya pada pertandingan playoff [12]. Data diambil dari kuesioner yang dibagikan kepada partisipan, yang berisi kuesioner berupa nama tim pemenang yang dipilih, tingkat kepercayaan atas pilihan tim, dan berapa persen kira-kira tim tersebut menang. Kuesioner ini mengacu kepada pertandingan playoff babak pertama, dari 15 pertandingan hanya 1 yang gagal diprediksi dan jika dibandingkan dengan 3 data source yang lainnya, kesalahan yang terjadi dengan metode ini adalah 1/15. Metode yang digunakan akan berjalan dengan baik jika partisipan dalam mengisi kuesionernya memiliki wawasan yang baik akan olahraga basket NBA.

Selain menggunakan Surprisingly Popular Algorithm, peneliti lain menggunakan Linear Regression untuk menentukan probabilitas dari hasil pertandingan di antara dua tim di NBA serta distribusi binomial untuk menentukan banyaknya pertandingan yang akan dimenangkan antara dua tim tersebut[13], dengan menggunakan data dari 5 regular season dari 2014-2019, data yang digunakan adalah 10 parameter yang sering digunakan dalam statistik bola basket dan terdapat tiga cara pembagian dalam data testing dan training. Hasil koreksi sebesar 49% dari hasil prediksi tanpa kesalahan serta 92% untuk game yang membutuhkan 1 pertandingan lagi. Pada penelitian ini juga menemukan bahwa shoot rates dan banyaknya rebound merupakan faktor utama.

Penelitian yang dilakukan Wang et al [14] prediksi pemenang pertandingan playoff menggunakan decision tree dan linear regression berdasarkan data yang

diambil dari reference basketball dari tahun 2009-2019[14]. Data yang digunakan dibagi menjadi 2 dataset, yaitu data asli dan data yang sudah di pre-processing sebelum digunakan. Hasil yang didapat dari hasil penelitian menunjukkan bahwa penggantian metode decision tree/linear regression tidak berdampak signifikan, yang berpengaruh signifikan adalah dataset yang digunakan serta pemilihan acak yang dilakukan.

Pemodelan dengan pendekatan yang baru dengan memperhitungkan hubungan antar pemain dalam pertandingan untuk menentukan performa tim [15]. Dengan menggunakan dataset pertandingan tahun 2014-2015 dan menggunakan model Linear Regression, Neural Network, serta Quadratic Model dalam menentukan performa tim. Hasil yang menyatakan bahwa ketiga model ini menghasilkan performa yang bagus. Model quadratic lebih stabil dibandingkan 2 model yang lainnya dan model ini mampu untuk memeriksa sinergi dan adversity dari pemain di suatu tim.

Selain metode machine learning, algoritma PageRank yang biasanya digunakan untuk menentukan peringkat halaman website di search engine google untuk menentukan peringkat dari tim NBA [16]. Paper ini membuktikan bahwa model dari PageRank mencerminkan keadaan dari keseluruhan musim dan hasil yang digunakan hampir sama dengan peringkat yang sebenarnya dalam playoffs. Data yang digunakan adalah data pada tahun 2018-2019, terdapat beberapa masalah dalam memprediksi hasilnya seperti pemain yang cedera, gaya bermain, dan hal-hal yang terjadi selama musim tersebut yang tidak dapat diprediksi.

Penelitian oleh Z. Mahmood et al. [17] menggunakan machine learning untuk menentukan apakah pemain tersebut adalah pemain bintang atau pemain bukan bintang. Penemuan pemain bintang merupakan hal yang penting dalam sebuah tim NBA. Selain bisa membantu tim untuk mendapatkan gelar juara atau memenangkan suatu pertandingan, pemain bintang juga bisa untuk menaikkan pendapatan dari sebuah tim, dari penjualan merchandise dll. Penemuan pemain bintang ini juga digunakan agar penyusunan strategi untuk memenangkan pertandingan menjadi semakin efisien.

Penelitian ini menggunakan NB, MEM, CART, BN, dan SVM sebagai model dan dataset dari pertandingan season 2004-2009. Hasil dari penelitian ini membuktikan bahwa machine learning mampu untuk memprediksi seorang pemain bintang atau tidak, tetapi hanya pada musim pertandingan tersebut saja, tidak mampu untuk mendeteksi kemampuan pemain dimasa yang akan datang.

Penelitian oleh Wang et al. [18] menggunakan *neural network* untuk menentukan *play call* dari tim yang diamati. Data diambil dari SportVU *dataset* yang untuk tim Toronto Raptors. Hasil percobaan dari Wang et al. membuktikan bahwa metode *neural network* mampu bekerja dengan baik dalam memprediksi *play call* dari tim yang diamati. Penggunaan model RNN yang disesuaikan dan representasi data yang lebih baik meningkatkan kinerja dalam memprediksi *play call* lebih tinggi lagi. Hasil penelitian ini mampu membuat pengguna untuk menentukan strategi yang lebih baik dalam pertandingan untuk tim dan mampu mengetahui pola pergerakan dari tim lawan.

Penelitian oleh Hore et al [19] menggunakan *machine learning* untuk menentukan pemain yang akan tetap bertahan bermain di liga NBA lebih dari 5 tahun. Data yang digunakan terdiri dari 1300 data yang terdiri dari 20 parameter yang nantinya akan *preprocessing* menjadi 11 parameter. Penelitian ini menggunakan empat algoritma seperti *Naïve Bayes*, *Support Vector Machine*, *Multilayer Feed Forward Network*, dan *Random Forest*. Hasil dari penelitian ini berdasarkan akurasi dan statistik Kappa serta kurva ROC dan uji McNemar menunjukkan bahwa SVM memiliki performa terbaik.

Penelitian oleh Lu et al [20] menggunakan *machine learning* untuk membangun model yang sederhana dan menentukan skor dari pertandingan NBA. Lu menggunakan data dari musim tahun 2012- 2016. Model yang bekerja secara bagus untuk memprediksi hasil dari pertandingan NBA, dimana hasil ini hampir sama dengan hasil yang dimiliki oleh bandar judi untuk pertandingan NBA. Model yang dihasilkan

pada paper ini sederhana dan mampu bekerja dengan baik untuk memprediksi pertandingan pada musim 2017-2018.

Penelitian oleh Zovak *et al* [21] menggunakan *machine learning* untuk memprediksi performa pemain berdasarkan data statistik dari pemain selama musim-musim sebelumnya dan sekarang. Zovak menggunakan data dari pemain NBA pada musim 2017-2018 dan *subset* pemain yang dipilih adalah Steph Curry. Zovak *et al* menggunakan model SVM dan Linear Regression dalam pembuatan modelnya. Hasil penelitian ini adalah model mampu untuk memprediksi kemampuan pemain Y dalam suatu pertandingan X (pertandingan yang dipilih), apakah *underated* atau *overated*.

Penelitian oleh Migliorati *et al* [22] menggunakan *machine learning* untuk menentukan faktor yang akan mendukung suatu tim menuju kemenangan. Migliorati menggunakan data dari liga NBA musim 2004-2005 sampai 2017-2018 khususnya data Golden State Warriors. Model *machine learning* yang digunakan adalah CART dan Random Forest. Hasil dari penelitian ini bahwa *machine learning* mampu untuk mendeteksi faktor-faktor penentu kemenangan dari suatu tim, pada kasus *paper* ini faktor terpenting untuk megnantarkan Golden State Warrior untuk menang dalam suatu pertandingan adalah faktor *shooting*, setelah itu lalu faktor *defending* (*defensive rebounds* dan *opponent turnovers*).

Penelitian oleh Chou *et al* [23] menggunakan *machine learning* untuk menentukan pemain yang akan menjadi *Hall of Fame* (julukan untuk seseorang yang sangat berprestasi dan berperan besar dalam liga NBA). Data yang digunakan sebanyak 4728 data statistik dan penghargaan para pemain Sampe diambil dari 85 pemain yang dinyatakan sebagai *Hall of Fame* dan 113 pemain yang *non-Hall of Fame*, berdasarkan data lengkap dan panjang karir yang lama (≥ 15 tahun). Chou *et al* menggunakan metode ANN dan CNN dalam pembuatan model. Hasil penelitian ini adalah model ANN mengungguli model CNN, 20 variabel fitur dalam ANN menghasilkan AUC lebih tinggi, dan aplikasi untuk memprediksi *Hall of Fame* berhasil dikembangkan.

Penelitian oleh Hsu *et al* [24] menggunakan *machine learning* untuk memprediksi *rank* untuk setiap tim yang terdapat didalam liga NBA. Data yang digunakan adalah data semua tim dari liga NBA musim 2011-2018. Penelitian ini menggunakan metode Support Vector Machine, Polynomial Regression, dan Random Forest dalam menciptakan model. Hasil dari penelitian ini adalah dari ketiga model tersebut, Support Vector Machine menghasilkan hasil yang lebih baik dalam memprediksi *rank* untuk setiap tim di liga NBA.

Penelitian oleh Chi *et al* [25] menggunakan pembobotan adaptif yang diintegrasikan dengan metode *machine learning* untuk memprediksi hasil pertandingan NBA. Penelitian ini menggunakan lima metode yaitu Classification and Regression Trees (CART), Random Forest (RF), Stochastic Gradient Boosting (SGB), eXtreme Gradient Boosting (XGBoost), and Extreme Learning Machine (ELM). Data yang digunakan pada penelitian ini adalah data pertandingan pada liga NBA musim 2018-2019. Hasil dari penelitian ini adalah bahwa *machine learning* dapat memprediksi pertandingan NBA dengan baik, jika ditambah dengan pembobotan adaptif untuk datanya maka akan menghasilkan performa yang lebih baik lagi.

Penelitian oleh Horvat *et al* [26] menggunakan *machine learning* dan metode validasi silang untuk memprediksi hasil pertandingan NBA. Penelitian ini menggunakan 7 metode yaitu Logistic Regression, Naïve Bayes, Decision Trees, Multilayer Perceptron Neural Network, Random Forest, K-NN, dan LogitBoost. Data yang digunakan adalah data dari liga NBA musim 2009/2010 sampai 2017/2018. Hasil penelitian ini adalah model terbaik yang didapatkan adalah dengan menggunakan Nearest Neighbors sedangkan model terburuk adalah Decision Tree, selain itu validasi dengan menggunakan validasi silang menghasilkan performa yang lebih baik dibandingkan menggunakan metode validasi Train dan Test.

Setelah melihat beberapa penelitian yang dilakukan, banyak yang menggunakan metode *machine learning* dalam melakukan penelitian dengan hasil yang

baik pada tema NBA. Penelitian ini membandingkan performa dari metode *machine learning* yang sering digunakan seperti ANN, GNB, SVM, dan RF. Penelitian ini ingin menganalisa lebih jauh mengapa banyak peneliti yang menggunakan *machine learning* dan metode apa yang terbaik diantara metode-metode tersebut untuk kasus ini.



Tabel 2. Perbandingan dengan penelitian sebelumnya

No	Pembanding/Peneliti	Algoritma	Pokok Penelitian	Data	Hasil
1	Lee, Michael D, Vi, Julie Danileiko, Irina [12]	Suprisingly Popular	Prediksi hasil pertandingan playoff	Kuesioner	Dari 15 pertandingan babak pertama hanya 1 yang salah prediksi dan mampu bersaing dengan 3 data source yang sering dipakai.
2	He, Wei-de, Chiang, Yi-ting[13]	Linear Regression dan sdistribusi binomial	Prediksi pertandingan tim NBA	Data dari 5 regular season dari 2014-2019	Hasil koreksi sebesar 49% dari hasil prediksti tanpa kesalahan serta 92% untuk game yang membutuhkan 1 pertandingan lagi. Pada penelitian ini juga menemukan bahwa shoot rates dan banyaknya rebound merupakan faktor utama.
3	Wang, Jingru Fan, Qishi [14]	Decision Tree dan Linear Resgesion	Prediksi pemenang pertandingan playoff	Data dari season 2009-2019	Penggantian metode decision tree/linear regression tidak berdampak signifikan, yang berpengaruh signifikan adalah dataset yang

					digunakan serta pemilihan acak yang dilakukan
4	Singh, Prastuti, Wang, Bai Yang[15]	Linear Regression, Neural Network, dan Quadratic Model	Prediksi pertandingan berdasarkan hubungan antar pemain	Dataset pertandingan tahun 2014-2015	Hasil yang menyatakan bahwa ketiga model ini menghasilkan performa yang bagus. Model quadratic lebih stabil dibandingkan 2 model yang lainnya dan model ini mampu untuk memeriksa sinergi dan adversity dari pemain di suatu tim
5	Yang, Fan,Zhang, Jun[16]	PageRank	Prediksi peringkat NBA playoff	Data 2018-2019	Paper ini membuktikan bahwa model dari PageRank mencerminkan keadaan dari keseluruhan musim dan hasil yang digunakan hampir sama dengan peringkat yang sebenarnya dalam playoffs

6	Z. Mahmood, A. Daud, and R. A. Abbasi [17]	SVM,CART,Baynesian, Maximum Entropy Markov Model (MEM),Naïve Bayes	Prediksi Bintang Baru Pemain NBA	Data 5 season 2004-2009	Paper ini membuktikan bahwa machine learning mampu untuk menilai apakah pemain ini adalah pemain bintang atau tidak
7	Wang, Kuan-chieh Zemel, Richard[18]	NN, RNN	Prediksi <i>play call</i> dalam pertandingan basket	Data 2 season 2013-2015	Paper ini membuktikan bahwa machine learning mampu untuk memprediksi pola <i>play call</i> yang digunakan untuk dalam strategi basket selama pertandingan
8	Hore, Sirshendu Bhattacharya, Tanmay [19]	Naïve Bayes, Support Vector Machine, Multilayer Feed Forward Network, dan Random Forest	Prediksi pemain yang akan bertahan di liga NBA selama 5 tahun atau lebih	Data statistik pemain dengan jumlah 1300 data dengan 20 parameter	Paper ini membuktikan bahwa machine learning mampu memprediksi pemain yang akan bertahan di liga NBA selama 5 tahun atau lebih dan model terbaik pada penelitian ini adalah SVM
9	Lu, Jiayi Chen, Yuan Zhu, Yida[20]	Linear Regression (Least Square Estimation)	Prediksi skor pertandingan NBA	Data dari musim 2012-2016	Paper ini membuktikan bahwa machine learning

					mampu memprediksi skor pertandingan liga NBA
10	Zovak, Trpimir Šarcevic, Ana Vranic, Mihaela Pintar, Damir[21]	SVM dan Linear Regression	Prediksi kemampuan pemain dalam suatu pertandingan	Data para pemain pada tahun 2017-2018	Paper ini membuktikan bahwa machine learning mampu untuk memprediksi kemampuan pemain dalam suatu pertandingan
11	Migliorati, Manlio [22]	CART dan Random Forests	Prediksi faktor terpenting dalam suatu tim di NBA untuk memenangkan pertandingan	Data Golden State Warrior musim 2004-2005 sampai 2017-2018	Paper ini membuktikan bahwa machine learning mampu untuk menentukan faktor terpenting dari suatu tim untuk mencapai kemenangan dan membantu meningkatkan kemampuan tim
12	Chou, Po Hsin Chien, Tsair Wei Yang, Ting Ya Yeh, Yu Tsen Chou, Willy Yeh, Chao Hung [23]	ANN dan CNN	Prediksi pemain yang akan menjadi Hall of Fame	Data statistik sejumlah 4728 dari 85 pemain Hall of Fame dan 113 pemain non Hall of Fame	Paper ini membuktikan bahwa machine learning mampu untuk memprediksi pemain yang akan menjadi Hall of Fame

					dan menemukan bahwa ANN lebih baik dari CNN untuk kasus ini, selain itu aplikasi berhasil dibangun
13	Hsu, Po Han Galsanbadam, Sainzaya Yang, Jr Syu Yang, Chan Yun[24]	SVM, Polynomial Regression, dan Random Forest	Prediksi rank setiap tim di liga NBA	Data statistik setiap tim dari musim 2011-2018	Paper ini membuktikan bahwa machine learning mampu untuk memprediksi rank dari setiap tim di liga NBA berdasarkan data statistik
14	Chi, Tian, Chien, Wei[25]	CART, RF, SGB, XGBoost, dan ELM	Prediksi pertandingan NBA	Data season 2018-2019	Paper ini membuktikan bahwa machine learning mampu untuk memprediksi hasil pertandingan NBA dan menemukan bahwa mengintegrasikan dengan pembobotan adaptif terhadap data, membuat performa machine learning menjadi lebih baik

15	Horvat, Tomislav Havaš, Ladislav Srpak, Dunja[26]	Logistic Regression, Naïve Bayes, Decision Threes, Multilayer Perceptron Neural Network, Random Forest, K-NN, dan LogitBoost	Prediksi pertandingan NBA	Data season 2009/2010 sampai 2017/2019	Paper ini membuktikan bahwa machine learning mampu untuk memprediksi hasil pertandingan NBA dan pada paper ini model terbaik dengan menggunakan metode Nearest neighbors, sedangkan model terburuk dengan menggunakan metode Decision Threes, menggunakan metode validasi silang menghasilkan performa yang lebih baik
16	Nusa Satria et al	RF, NB, SVM, ANN	Prediksi pertandingan regular season	Data 5 season 2015-2020	Pembuktian mengenai performa antara metode machine learning dan stretegi untuk mencapai kemenangan tim Bucks