

BAB II

KAJIAN PUSTAKA

2.1. Penelitian Terkait

Penelitian ini menggunakan beberapa acuan terkait penelitian sebelumnya yang berkaitan dengan metode *K-Nearest Neighbors*. Adapun beberapa penelitian terkait yang telah dilakukan sebelumnya adalah sebagai berikut:

Jurnal dengan judul **Penerapan Algoritma *K-Nearest Neighbor* (K-NN) Dengan Pencarian Optimal Untuk Prediksi Prestasi Siswa** (Umaidah & Purwantaro, 2019). Penelitian ini dilatar belakangi oleh masih banyaknya siswa yang merasakan permasalahan dalam proses pembelajaran. Diantaranya adalah siswa yang mendapatkan kesempatan belajar yang baik, kemampuan yang cukup baik, namun hasil yang dicapai justru menunjukkan bahwa anak tersebut tidak memiliki prestasi. Pada penelitian ini menunjukkan bahwa terdapat hambatan dan masalah dalam proses pencapaian prestasi siswa itu sendiri. Sehingga dibutuhkan suatu sistem untuk mengetahui prestasi siswa tersebut.

Jurnal dengan judul **Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *K-Nearest Neighbour*** (Yandi Saputra & Primadasa, 2018). Penelitian ini dilatar belakangi oleh peningkatan jumlah siswa setiap tahunnya yang menyebabkan penumpukan data mahasiswa. Sehingga dibutuhkan suatu sistem prediksi ketepatan kelulusan mahasiswa dirancang untuk mendukung prodi untuk membimbing mahasiswa agar kelulusan tepat waktu. Menurut peneliti, dengan mengetahui prediksi status kelulusan mahasiswa selama masa perkuliahan, maka prodi yang dibantu oleh dosen pembimbing akademik dapat memberikan perhatian khusus kepada mahasiswa yang diprediksikan tidak lulus tepat waktu (terlambat) agar mahasiswa tersebut dapat meningkatkan indeks prestasi mereka setiap semester sehingga mereka dapat lulus tepat waktu. Penelitian ini

menggunakan 9 data latih dan 1 data uji. Dengan menggunakan $k=5$ yang diterapkan menggunakan metode *K-Nearest Neighbor* untuk memprediksi kelulusan siswa.

Jurnal dengan judul **Prediksi Potensi Siswa Putus Sekolah Akibat Pandemi Covid-19 Menggunakan Algoritma *K-Nearest Neighbor*** (Darmayanti et al., 2021). Pandemi Covid-19 yang menyebar di dunia khususnya di Indonesia berdampak pada berbagai bidang. Salah satunya pendidikan yaitu tentang ancaman siswa putus sekolah. Putus sekolah sangat berdampak negatif bagi siswa sehingga mengalami penurunan kepercayaan diri dan motivasi untuk belajar. Prediksi secara dini harus dilakukan, untuk mencegah banyaknya siswa yang putus sekolah, salah satunya menggunakan teknik *data mining*. *Dataset* yang digunakan dalam penelitian ini diambil dari siswa di Sekolah Menengah Pertama (SMP) yang ada di Kabupaten Banyumas. Metode yang digunakan dalam penelitian ini yaitu *confusion matrix* dan *10-fold cross validation* pada algoritme *K-Nearest Neighbor* (KNN). Hasil yang didapatkan pada algoritme KNN dalam mempredksi potensi siswa putus sekolah sebesar 87.4214%, dengan nilai *precision* 88.2%, *recall* 87.4% dan *F-Measure* 87%. Maka hasil nilai akurasi pada algoritme KNN berkategori *Good Classification* dan diperoleh hasil algortime yang efektif digunakan dalam memprediksi potensi siswa putus sekolah yaitu algoritme *K-Nearest Neighbor* (K-NN).

Jurnal dengan judul **Prediksi PENERAPAN ALGORITMA KLASIFIKASI K-NEAREST NEIGHBOR (K-NN) UNTUK MEMPREDIKSI UJIAN NASIONAL DENGAN MENGGUNAKAN HASIL NILAI TRY OUT SISWA SEKOLAH MENENGAH KEJURUAN BERBASIS DESKTOP PADA SMK BINA INFORMATIKA BINTARO.**(Aljabar & Kusumaningsih, 2018) penelitian ini dilatar belakang oleh kekhawatiran guru terhadap kesiapan siswa dalam menghadapi Ujian Nasional (UN). Oleh karena itu diselenggarakanlah ujian Hasil Try Out. Yang mana ujian tersebut dapat digunakan oleh siswa untuk mengetahui materi apa

yang sudah dikuasai dan yang belum dikuasai. Dari hasil tersebut diharapkan siswa mampu mengejar ketertinggalan terhadap materi yang belum dikuasai. Berdasarkan hasil ujian Try Out pihak sekolah menentukan prediksi apakah siswa/i SMK Bina Informatika Bintaro akan lulus atau tidak dalam Ujian Nasional (UN) berdasarkan nilai Try Out yang telah dicapai tersebut.

Dari keempat hasil penelitian tersebut dapat disimpulkan bahwa adanya persamaan dan perbedaan dalam penelitian ini. Perbedaan dari penelitian ini dan penelitian sebelumnya adalah penelitian pertama dan keempat menghasilkan sebuah aplikasi. Aplikasi yang diperoleh pada penelitian pertama adalah aplikasi untuk memprediksi ketepatan kelulusan mahasiswa dirancang untuk mendukung prodi untuk membimbing mahasiswa agar lulus tepat waktu menggunakan metode Data Mining yaitu Algoritma *K-Nearest Neighbor*, sedangkan aplikasi yang diperoleh dari penelitian keempat adalah aplikasi untuk memprediksi kelulusan siswa dengan menggunakan nilai Try Out sebagai acuannya. Pada penelitian kedua ketiga tidak menghasilkan aplikasi. Adapun persamaan antara penelitian pertama dan ketiga terletak pada metode yang digunakan oleh peneliti, pertama, ketiga dan keempat yaitu menggunakan metode *K-Nearest Neighbor* (K-NN) untuk menganalisis masalah yang diteliti oleh para peneliti dibidang masing-masing. Algoritma *K-nearest neighbors* atau KNN adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data yang sudah terklasifikasi sebelumnya yang diambil dari k tetangga terdekatnya. Berbeda dengan penelitian sebelumnya, pada penelitian ini akan dilakukan suatu pencarian nilai k yang optimal dengan metode *K-Nearest Neighbor* untuk menentukan hasil akurasi, *precision*, *recall* dan *f1-score* terbaik untuk memprediksi perkembangan belajar santri.

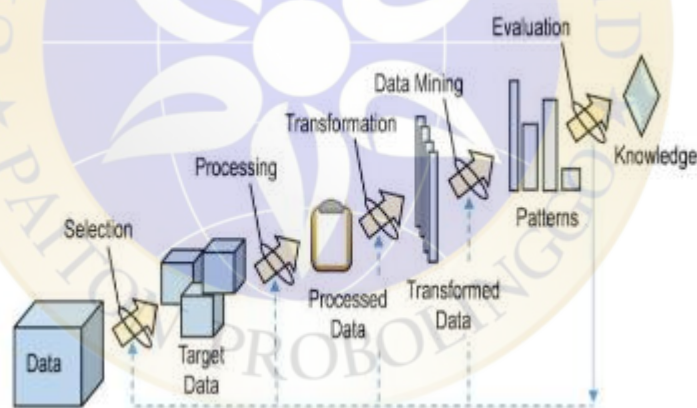
2.2. Landasan Teori

Pada bab ini akan dibahas tentang dasar teori mengenai istilah yang terdapat dalam penelitian, adapun dasar teori penelitian antara lain sebagai berikut:

2.2.1 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan dari sekumpulan data yang sangat besar. Data mining juga merupakan suatu langkah dalam *knowledge discovery in database* (KDD). (Mustafa & Simpen, n.d.)

Langkah-langkah dalam menemukan pengetahuan (*discovery knowledge*) pada gambar dibawah ini:



Gambar 2.1. Proses KDD

Secara keseluruhan, proses *Knowledge Discovery in Databases* (KDD) dapat digambarkan sebagai berikut:

1. *Data Selection* (pemilihan data)

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi di KDD dimulai. Data yang terpilih akan digunakan untuk proses data mining.

2. *PreProcessing Data* (pembersihan data)

Sebelum proses data mining dapat dilakukan, maka perlu dilakukan proses cleaning terlebih dahulu pada data yang menjadi fokus KDD. Proses pembersihan meliputi antara lain menghapus data duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data.

3. Transformasi (pengubah data)

Transformasi adalah mengubah data menjadi bentuk yang cocok untuk penambangan.

4. Penambangan Data (*Data Mining*)

Data mining adalah proses mencari pola atau informasi yang menarik pada data yang dipilih dengan menggunakan teknik atau metode tertentu.

5. Interpretasi/Evaluasi

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dipahami oleh pihak yang berkepentingan. Seperti menggunakan visualisasi atau tampilan yang dapat menjelaskan keluaran sistem.

2.2.1. Algoritma K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (K-NN) adalah metode yang digunakan untuk melakukan klasifikasi terhadap objek baru berdasarkan (K) tetangga terdekatnya. K-NN adalah algoritma pembelajaran *supervised learning*, di mana hasil query baru diklasifikasikan menurut mayoritas kategori di K-NN. Maka kelas yang paling banyak muncul adalah kelas hasil klasifikasi. *Nearest Neighbor* adalah metode pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama, dimana klasifikasi algoritma ini menggunakan voting terbanyak dari obyek k dan menggunakan ketetanggaan sebagai nilai prediksi. (Mustafa & Simpen, n.d.)

Untuk mendefinisikan jarak antara dua titik yaitu titik pada data training dan data testing maka digunakan rumus *Euclidean* seperti yang ditunjukkan pada persamaan dibawah ini

$$d(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$$

Dengan d adalah jarak antara titik pada data training x dan data training y yang akan diklasifikasi. (Munanda, 2020)

Langkah-langkah *K-Nearest Neighbor* sebagai berikut (Nurul et al., 2021)

1. Menentukan parameter k (jumlah tetangga paling dekat)
2. Menghitung kuadrat jarak *Euclidean* antara data uji baru yang akan dievaluasi
3. Mengurutkan jarak *Euclidean* dari yang terkecil
4. Mengumpulkan kategori y berdasarkan nilai k
5. Mencari jumlah kelas yang sering muncul sebagai hasil prediksi.

2.2.2. Prediksi

Menurut Kamus Besar Bahasa Indonesia, prediksi adalah hasil dari kegiatan meramal atau meramalkan ramalan atau ramalan. arti prediksi dalam istilah akan tergantung pada konteks atau masalah. Berbeda dengan arti ramalan dalam bahasa yang artinya prediksi atau perkiraan yang sudah menjadi pemahaman baku. Prediksi digunakan untuk memperkirakan nilai masa depan, misalnya memprediksi stok barang satu tahun ke depan. Fungsi ini meliputi *Neural Network*, *Decision Tree*, dan *Nearest*. Prediksi menggunakan beberapa variabel database atau bidang untuk memprediksi variabel nilai masa depan yang diperlukan, yang saat ini tidak diketahui. (Wiyata Mandala & Putri, 2018).

2.2.3. Perkembangan Belajar Santri

Dalam bahasa itu sendiri, perkembangan merupakan suatu tahapan, baik fisik maupun psikis, menuju tahapan berikutnya. Namun secara istilah merupakan tahapan yang berkesinambungan dan saling berhubungan. Sedangkan menurut Slameto (1995) belajar adalah proses bisnis yang dilakukan oleh individu untuk memperoleh perubahan tingkah laku secara keseluruhan sebagai hasil dari pengalaman individu dalam berinteraksi dengan lingkungannya.

Jadi dapat disimpulkan bahwa perkembangan belajar merupakan tahap perubahan yang terjadi baik dari fungsi, pola pikir, moral, fisik, maupun psikis menuju tahap selanjutnya yang saling berkelanjutan. Menurut Mansur Hidayat, santri di masyarakat pedesaan Jawa adalah komunitas Muslim yang taat beragama. Rizki menjelaskan ada pendapat yang menjadi acuan, pertama santri yang artinya melek huruf dalam bahasa setidaknya ada dua pendapat yang dapat digunakan sebagai referensi. Pertama, santri berasal dari kata "santri" dari bahasa Sansekerta yang artinya melek huruf. Kedua, kata santri yang berasal

dari bahasa Jawa "cantrik" yang artinya seseorang yang mengikuti seorang guru kemanapun pergi atau tinggal bersama tujuannya adalah untuk belajar ilmu padanya. Definisi ini sejalan dengan pemahaman siswa secara umum, yaitu: orang yang mempelajari Islam dan jelajahi Islam di pesantren yang menjadi tempat belajar bagi siswa. Jika ditelusuri ke tradisi pesantren, ada dua kelompok santri, yaitu *santri mukim*, santri dari daerah yang jauh dan menetap di asrama, dan biasanya telah mengambil tanggung jawab mengurus kepentingan sehari-hari pesantren, seperti mengajar santri tentang kitab tingkat rendah dan menengah; *santri kalong*, adalah santri yang berasal dari desa lingkungan sekitar, yang biasanya tidak tinggal di pesantren kecuali di waktu-waktu belajar saja. (Wahidin et al., 2021)

2.2.4. Lembaga Tamhidiyah

Sesuai dengan penamaan Tamhidiyah artinya persiapan, mahasiswi santri baru yang akan menjadi pengurus di daerah-daerah dipersiapkan terlebih dahulu dengan dibekali beberapa ilmu, diantaranya Al-qur'an dan Furudul 'Aininyah. Mahasiswi santri diasramakan di daerah Al Maziyah selama 1 tahun, furudul 'ainiyah juga merupakan salah satu pembelajaran yang harus dikuasai oleh mahasiswi santri Tamhidiyah, mengingat Mahasiswi Santri Tamhidiyah mayoritas santri baru, walaupun ada sebagian yang menempuh pendidikan SLTA di Nurul Jadid, yang mayoritas belum faham tentang bagaimana mempraktikkan Furudul 'aininyah mereka dalam kehidupan sehari-hari, sehingga penting bagi mereka untuk menguasai materi-materi yang sudah ditetapkan dan harus menguasai keseluruhan materi agar bisa lulus dari lembaga Tamhidiyah.