Studi kasus menggunakan dataset iris adalah tentang menganalisis dan memodelkan data untuk memprediksi spesies bunga iris berdasarkan fitur-fitur morfologisnya. Misalkan kita telah mengumpulkan data mengenai sepal length, sepal width, petal length, dan petal width dari berbagai spesimen bunga iris, serta label kelas yang menunjukkan spesies masing-masing bunga iris. Dengan menggunakan dataset ini, kita dapat mengimplementasikan berbagai algoritma pembelajaran mesin untuk membuat model yang dapat memprediksi spesies bunga iris baru berdasarkan fitur-fitur morfologisnya.

Tujuan utama dari dataset ini adalah untuk mengembangkan model prediktif yang dapat memprediksi spesies bunga iris berdasarkan fitur-fitur morfologisnya. Hal ini dapat bermanfaat bagi para peneliti, petani, atau pecinta tumbuhan untuk mengidentifikasi spesies bunga iris dengan lebih cepat dan akurat

Situasi bisnis yang mendasari analisa ini adalah industri pertanian botani yang saat ini proses identifikasi dilakukan secara manual oleh ahli botani, yang memerlukan waktu dan tenaga yang cukup besar

Tujuan data mining pada dataset ini adalah mengekstraksi informasi yang berharga dan wawasan yang mendalam tentang hubungan antara fitur-fitur morfologis bunga iris dan spesiesnya

Rencana proyek untuk menganalisa dataset ini adalah

1. Pemahaman Data

* Memahami struktur dataset iris

1. Eksplorasi data

* Melakukan eksplorasi data untuk memahami fitur-fitur separti sepallength dan lain lain
* Melakukan visualisasi data untuk mendapat wawasan tambahan

1. Pra-pemrosesan data
   * Memeriksa missing value (jika ada) dan menanganinya
   * Melakukan pemrosesan data
2. Pembagian data

* Membagi dataset menjadi set pelatihan dan set pengujian

1. Pemodelan

* Memilih model machine learning
* Melatih model menggunakan set pelatihan
* Menggunakan Teknik validasi silang jika perlu untuk mengevaluasi model secara lebih akurat

1. Evaluasi model

* Mengevaluasi kinerja model
* Membandingkan performa model yang berbeda jika lebih dari satu model yang digunakan

1. Tuning model (opsional)

* Jika diperlukan, melakukan penyetelan parameter (parameter tuning) pada model untuk meningkatkan kinerjanya.

1. Prediksi

* Menggunakan model yang telah ditrain untuk membuat prediksi spesies iris berdasarkan fitur-fitur morfologisnya pada data baru atau data pengujian

1. Interpretasi hasil

* Menginterpretasikan hasil prediksi dan mengevaluasi apakah model tersebut dapat digunakan secara efektif dalam memprediksi spesies iris dari fitur-fitur yang diberikan

Pada tahap awal, pemahaman data dilakukan dengan mempelajari struktur dataset iris, yang mencakup fitur-fitur seperti sepallength, sepalwidth, petallength, dan petalwidth. Setelah itu, dilanjutkan dengan eksplorasi data untuk mendapatkan pemahaman yang lebih dalam tentang fitur-fitur tersebut melalui visualisasi data. Selanjutnya, tahap pra-pemrosesan data dilakukan dengan memeriksa dan menangani missing value, serta melakukan pemrosesan data lainnya untuk mempersiapkan dataset. Pembagian data dilakukan dengan membagi dataset menjadi set pelatihan dan pengujian untuk melatih dan menguji model machine learning yang dipilih. Setelah model dipilih, dilakukan pelatihan menggunakan set pelatihan dan evaluasi kinerjanya, termasuk menggunakan teknik validasi silang jika diperlukan. Evaluasi model melibatkan perbandingan performa model yang berbeda jika digunakan lebih dari satu model, dengan opsi untuk menyetel parameter untuk meningkatkan kinerja model. Setelah model ditentukan, tahap prediksi dilakukan menggunakan model yang telah ditraining untuk membuat prediksi spesies iris berdasarkan fitur-fitur morfologisnya pada data baru atau data pengujian. Akhirnya, interpretasi hasil dilakukan untuk mengevaluasi efektivitas model dalam memprediksi spesies iris dari fitur-fitur yang diberikan.