**Laporan Penyisihan Data Mining (Iris DataSet)  
Arianto Wibowo – 5113100037**

**PreProcess**  
Data yang dimiliki pertama dibagi dua menjadi dahulu, yaitu data untuk training dan data untuk testing. DataSet Iris memiliki 150 record, sehingga data ini dibagi menjadi 140 data training dan 10 data testing secara random. Untuk 140 data training disimpan ke file bernama IrisTrain.txt dan 10 data testing disimpan ke file bernama IrisTest.txt. Selain itu, kedua data ini belum memiliki nama atributnya. Oleh karena itu, ditambahkan 1 row bagian paling atas (header) yang berisi kelas dari atribut tersebut. Atribut diambil dari deskripsi data set, yaitu (panjang sepal, lebar sepal, panjang petal, lebar petal, dan kelas iris). Disini , kelas iris terbagi menjadi 3, yaitu Iris-setosa, Iris-virginica, dan Iris-versicolor.

**Metode Klasifikasi**  
Untuk metode klasifikasi, digunakan metode SVM (Support Vector Machine) untuk melakukan proses klasifikasi terhadap data iris tersebut. SVM dipilih karena akurasinya yang cukup bagus dan kompleksitasnya yang tidak terlalu besar. Karena dataset yang hanya berjumlah 140 record, maka SVM seharusnya dapat melakukan klasifikasi dengan cukup cepat karena kompleksitasnya worst casenya O(n3) dimana n adalah jumlah training data (140). Setelah proses training selesai, maka program akan di test menggunakan 10 data testing yang telah disiapkan sebelumnya di IrisTest.txt. Dari data test tersebut, output dikeluarkan ke satu file baru yaitu answer.txt. Output yang dikeluarkan ke answer.txt adalah kelas asli yang berada pada data test dan hasil prediksinya menggunakan SVM. Oleh karena itu, terdapat 10 row dan 2 kolom pada answer.txt dimana 1 row merepresentasikan 1 record pada IrisTest.txt dan kolom pertama merupakan kelas asli iris tersebut dan kolom kedua adalah hasil prediksi klasifikasi tersebut. Setelah itu, juga ditunjukkan akurasi dari hasil klasifikasi, yaitu jumlah prediksi benar dibagi jumlah data. Terakhir, training time juga dioutputkan sebagai satuan ukuran besarnya waktu yang dibutuhkan untuk melakukan training terhadap SVM tersebut.  
  
**Tools**  
Untuk melakukan proses klasifikasi tersebut, tools yang digunakan adalah python dengan numpy dan scipy. Untuk IDE, digunakan spyder yang memudahkan proses coding di python. Numpy dan Scipy ini (beserta scikit learn) sudah berada dalam satu paket di anaconda, sehingga cukup menginstall anaconda untuk menjalankan program ini.  
  
**Output**  
Output dari hasil klasifikasi pada Iris Dataset ternyata sangat baik. Dari 10 data testing yang diambil secara acak, ternyata SVM berhasil memprediksi seluruh data tersebut sehingga akurasinya 100%. Selain itu, waktu yang dibutuhkan untuk pembelajaran juga relatif kecil, yaitu sekitar 0.001 sekon. Hasil ini dapat dilihat sepenuhnya di file answer.txt.