**NAMA: ALRISKY LISMANTO**

**KELAS: 3KA29**

**NPM: 10121099**

**TUGAS 1 KONSEP DATA MINING**

1. **Data yang digunakan**: Data yang digunakan dalam penelitian ini adalah data siswa yang terdiri dari 416 siswa tingkat X SMK Negeri 4 Surakarta. Artikel ini menggunakan tipe data **Decision Tree** untuk melihat faktor-faktor probabilitas yang akan mempengaruhi alternatif-alternatif prestasi belajar siswa, disertai dengan prediksi hasil akhir yang akan didapat jika faktor-faktor Decision Tree terpenuhi. Menggunakan Atribut Diskrit (Kualitatif), serta menggunakan dimensi Regresi linier untuk membentuk model hubungan antara variabel terikat (dependen) dengan satu atau lebih variabel bebas (independen). Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai regresi linier sederhana, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai regresi linier berganda.
2. **Preprocessing**: Preprocessing yang dilakukan pada data meliputi pengumpulan data untuk menghasilkan data mentah (raw data) yang dibutuhkan oleh data mining, yang kemudian dilanjutkan dengan transformasi data untuk mengubah data mentah menjadi format yang dapat diproses oleh data mining, misalnya melalui filtrasi atau agregasi.
3. **Task Mining**: Penelitian ini bertujuan untuk membuat prediksi prestasi belajar siswa berdasarkan status sosial ekonomi orang tua, motivasi, kedisiplinan siswa dan prestasi masa lalu menggunakan metode data mining dengan algoritma J48. Data penelitian dianalisis juga dengan CHAID (Chi Squared Automatic Interaction Detection) dan regresi ganda.
4. **Hasil**: Hasil penelitian menunjukkan bahwa analisis prediksi menggunakan decision tree algoritma J48 memiliki akurasi sebesar 95,7%, sedangkan analisis prediksi menggunakan CHAID memiliki tingkat akurasi 82,1% dan analisis regresi ganda menghasilkan tingkat signifikansi sebesar 90,6%.
5. **Tools**: Analisis data yang pertama adalah dengan menggunakan teknik decision tree algoritma J48. Software yang digunakan adalah WEKA 3.6.9, lalu analisis yang kedua menggunakan CHAID (Chi Squared Automatic Interaction Detection). Analisis ini menggunakan software SPSS 16.0. Analisis yang ketiga menggunakan regresi ganda. Analisis ini menggunakan software SPSS 16.0.