**Laporan Uts Pembelajaran Mesin**

*Foris Juniawan Hulu / 223400015*

**Laporan Analisis Hasil Model dan Justifikasi Pemilihan Model Terbaik**

Laporan ini bertujuan untuk membandingkan hasil analisis dari dua file tugas:

1. **File UTS No.1**: Berfokus pada eksplorasi awal data harga rumah di Jakarta Selatan.
2. **File UTS No.2**: Mengimplementasikan dan mengevaluasi model klasifikasi untuk dataset simulasi.

Laporan ini akan menganalisis hasil dari masing-masing file dan memberikan justifikasi pemilihan model terbaik berdasarkan metrik evaluasi.

**File UTS No.1: Eksplorasi Awal Harga Rumah**

**Konsep Utama:**

* **Memuat data**: Dataset harga rumah diambil dari file Excel melalui integrasi Google Drive.
* **Preprocessing**: Melibatkan normalisasi dan encoding variabel kategoris menggunakan Pandas dan Scikit-learn.
* **Analisis Model**:
  + Inisiasi regresi linier untuk prediksi harga rumah.
  + Model belum dievaluasi secara menyeluruh karena file ini lebih berfokus pada persiapan data.

**Kesimpulan Sementara:**

File ini bertindak sebagai langkah awal eksplorasi data dan persiapan model. Langkah berikutnya adalah melakukan pelatihan model regresi linier dan mengevaluasi hasilnya dengan metrik seperti Mean Squared Error (MSE) atau R-squared.

**File UTS No.2: Evaluasi Model Klasifikasi**

**Konsep Utama:**

* **Preprocessing**: Mengatasi nilai kosong, standarisasi, dan pembagian dataset menjadi train-test.
* **Model yang Dievaluasi**:
  + Logistic Regression
  + k-Nearest Neighbors (kNN)
  + Support Vector Machine (SVM)
* **Hasil Evaluasi**:

| **Model** | **Akurasi** | **Presisi** | **Recall** | **F1-Score** |
| --- | --- | --- | --- | --- |
| Logistic Regression | 87.5% | 86.14% | 88.78% | 87.44% |
| k-Nearest Neighbors (kNN) | 85.5% | 85.57% | 84.69% | 85.13% |
| Support Vector Machine (SVM) | 88.0% | 86.27% | 89.80% | 88.00% |

**Analisis Model:**

1. **Logistic Regression**: Performa mendekati SVM tetapi memiliki recall yang lebih rendah, sehingga kurang optimal dalam mengenali kelas positif.
2. **k-Nearest Neighbors (kNN)**: Performa paling rendah dengan akurasi dan F1-Score yang lebih rendah dibandingkan model lainnya.
3. **Support Vector Machine (SVM)**: Kinerja terbaik dengan nilai tertinggi pada akurasi (88%), recall (89.80%), dan F1-Score (88%).

**Kesimpulan dan Justifikasi:**

* **Model Terbaik**: Support Vector Machine (SVM).
* **Alasan Pemilihan**:
  + SVM memiliki keseimbangan terbaik antara akurasi, presisi, recall, dan F1-Score.
  + Recall yang tinggi menjadikannya pilihan unggul untuk tugas yang memprioritaskan pengenalan kelas positif.

**4. Kesimpulan Umum**

* **File UTS No.1**: Memfasilitasi eksplorasi awal data dan persiapan untuk analisis regresi.
* **File UTS No.2**: Menghasilkan analisis komprehensif terhadap model klasifikasi, dengan SVM dipilih sebagai model terbaik.

**5. Rekomendasi**

1. **Untuk UTS No.1**: Lanjutkan analisis regresi dengan fokus pada evaluasi hasil menggunakan MSE atau R-squared.
2. **Untuk UTS No.2**: Eksplorasi hyperparameter pada SVM untuk meningkatkan performa lebih lanjut.