

Resultados Y Conclusiones

En este informe, se presentan los resultados y análisis comparativos de diferentes modelos de clasificación aplicados a un conjunto de datos de precios de casas. Se evaluaron cuatro modelos: Random Forest y tres variantes de SVM con diferentes kernels (lineal, RBF y polinomial). Se utilizaron métricas de evaluación como la precisión (accuracy) y la matriz de confusión para analizar el rendimiento de cada modelo.

Metodología

Preprocesamiento de Datos: Se realizaron las siguientes etapas de preprocesamiento en los datos:

- Separación de características y variable objetivo.
- Manejo de valores nulos (en este caso, se llenaron con ceros).
- Codificación one-hot para convertir variables categóricas en numéricas.
- División de los datos en conjuntos de entrenamiento y prueba (80% entrenamiento, 20% prueba).
- Escalado de características utilizando StandardScaler.

Modelos de Clasificación Evaluados:

- Random Forest
- SVM con kernel lineal
- SVM con kernel RBF
- SVM con kernel polinomial

Resultados

1. **Random Forest:**

- Precisión (Accuracy): Aproximadamente 99.7%
- Matriz de Confusión: Mayoría de las muestras clasificadas correctamente en todas las categorías, con solo un error.

2. **SVM (Kernel Lineal):**

- Precisión (Accuracy): 100%
- Matriz de Confusión: Todas las muestras clasificadas correctamente, sin errores.

3. **SVM (Kernel RBF):**

- Precisión (Accuracy): Aproximadamente 96.6%
- Matriz de Confusión: La mayoría de las muestras clasificadas correctamente, con algunos errores, especialmente en la categoría "Expensive".

4. **SVM (Kernel Polinomial):**

- Precisión (Accuracy): Aproximadamente 68.8%
- Matriz de Confusión: Rendimiento inferior en comparación con otros modelos, con un mayor número de errores en todas las categorías.

Análisis Comparativo

- **Mejor Rendimiento:** El modelo Random Forest muestra la mejor precisión, seguido por SVM con kernel lineal y RBF.
- **SVM vs. Random Forest:** Aunque SVM con kernel lineal y RBF tienen un rendimiento cercano al modelo Random Forest, este último es más robusto y no requiere tanto ajuste de hiperparámetros.
- **Kernel Polinomial:** Este kernel mostró el peor rendimiento, sugiriendo que no es adecuado para este conjunto de datos o que necesita ajustes adicionales.
- **Conclusión:** En general, el modelo Random Forest es la mejor opción para este problema de clasificación debido a su alta precisión y capacidad para manejar eficazmente diferentes tipos de datos.

Recomendaciones

- Basado en los resultados, se recomienda utilizar el modelo Random Forest para futuras predicciones de precios de casas.
- Se pueden explorar ajustes adicionales en los hiperparámetros de los modelos SVM para mejorar su rendimiento, especialmente el kernel polinomial.
- Es importante continuar evaluando y ajustando los modelos a medida que se obtienen más datos o se modifican las características del problema.