

Detalle del modelo de Aprendizaje Automático que se ha desarrollado

Si bien se opta por un modelo, a continuación se explica el estudio que se hizo, los modelos intentados y el modelo final elegido:

Modelos Utilizados

1. Modelo de Regresión Lineal:

- **Descripción:** La regresión lineal ajusta una línea recta que minimiza la suma de los errores cuadráticos entre las predicciones y los valores reales.
- **Resultados:**
 - **Mean Squared Error:** 5.405034592019282e-25
 - **Mean Absolute Error:** 5.445911920654406e-13
 - **R2 Score:** 1.0
- **Conclusión:** Aunque el R^2 es perfecto, lo que sugiere un ajuste excelente, esto podría ser un indicativo de sobreajuste, especialmente si los datos de prueba son muy similares a los de entrenamiento.

2. Modelo de Ridge con Hiperparámetros Ajustados:

- **Descripción:** La regresión de Ridge añade regularización L2 a la regresión lineal, ayudando a prevenir el sobreajuste al penalizar grandes coeficientes.
- **Hiperparámetros Ajustados:**
 - **alpha:** 0.1
 - **solver:** auto
- **Resultados:**
 - **Mean Squared Error:** 9.108228739648518e-08
 - **Mean Absolute Error:** 0.00023938539023037582
 - **R2 Score:** 0.9999999999997289
- **Conclusión:** Este modelo ofrece un equilibrio entre precisión y prevención del sobreajuste gracias a su componente de regularización, lo que lo hace una excelente opción.

3. Modelo de Random Forest con Hiperparámetros Ajustados:

- **Descripción:** Random Forest es un algoritmo de ensamble que construye múltiples árboles de decisión y fusiona sus resultados para mejorar la precisión y controlar el sobreajuste.
- **Hiperparámetros Ajustados:**
 - **max_depth:** 20
 - **max_features:** 'sqrt'
 - **min_samples_leaf:** 1
 - **min_samples_split:** 2
 - **n_estimators:** 200

- **Resultados:**
 - **Mean Squared Error:** 1547.5628260718418
 - **Mean Absolute Error:** 30.87690448275861
 - **R2 Score:** 0.9953933198246926
- **Conclusión:** Aunque este modelo también tiene un rendimiento excelente, su implementación mostró problemas con algunos ajustes de hiperparámetros.

Decisión Final

Se opta por el **Modelo de Ridge con Hiperparámetros Ajustados** debido a su rendimiento balanceado y su capacidad de regularización, lo que reduce el riesgo de sobreajuste. Este modelo ofrece una predicción precisa y es más robusto frente a la variabilidad en los datos.

Modelo de Ridge con Hiperparámetros Ajustados

1. Arquitectura del Modelo:

- **Ridge Regression:** La regresión de Ridge es una técnica de regularización que mejora la precisión del modelo de regresión lineal al reducir el problema de sobreajuste. Se añade un término de penalización a la función de coste, que está controlado por el hiperparámetro alpha.

2. Algoritmos Utilizados:

- **Regresión Lineal:** Utilizada como base, ajusta una línea recta que minimiza la suma de los errores cuadráticos.
- **Ridge Regression:** Añade regularización L2 a la regresión lineal para prevenir el sobreajuste.

3. Ajuste de Hiperparámetros:

- **GridSearchCV:** Herramienta utilizada para buscar de forma exhaustiva a través de un rango de hiperparámetros especificados. En este caso, se realizó la búsqueda para el hiperparámetro **alpha**.
- **Hiperparámetros Ajustados:**
 - **alpha:** 0.1
 - **solver:** auto

4. Proceso de Entrenamiento y Evaluación:

- **División de Datos:** Los datos fueron divididos en conjuntos de entrenamiento y prueba, con una proporción de 80% para entrenamiento y 20% para prueba.
- **Imputación de Valores Faltantes:** Se utilizó la estrategia de la mediana para imputar los valores faltantes, asegurando que no haya datos nulos en los conjuntos de entrenamiento y prueba.

- **Entrenamiento del Modelo:** El modelo de Ridge fue entrenado utilizando los datos de entrenamiento con los hiperparámetros ajustados.
- **Evaluación del Modelo:** Las métricas de evaluación incluyen el Mean Squared Error (MSE), Mean Absolute Error (MAE) y el R2 Score.

5. Resultados del Modelo:

- **Mejores Hiperparámetros:** {'alpha': 0.1, 'solver': 'auto'}
- **Mean Squared Error:** 9.108228739648518e-08
- **Mean Absolute Error:** 0.00023938539023037582
- **R2 Score:** 0.99999999999997289

Proceso del Código

```
# Ajuste de hiperparámetros usando GridSearchCV con Ridge
param_grid = {
    'alpha': [0.1, 1, 10, 100],
    'solver': ['auto', 'svd', 'cholesky', 'lsqr', 'sag']
}
grid_search = GridSearchCV(estimator=Ridge(), param_grid=param_grid, cv=5, scoring='r2',
n_jobs=-1)
grid_search.fit(X_train_imputed, y_train)

# Mejor modelo
best_model = grid_search.best_estimator_
best_model.fit(X_train_imputed, y_train)

# Predicciones y evaluación
y_pred = best_model.predict(X_test_imputed)
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Modelo de Ridge con Hiperparámetros Ajustados")
print("Mejores Hiperparámetros:", grid_search.best_params_)
print("Mean Squared Error:", mse)
print("Mean Absolute Error:", mae)
print("R2 Score:", r2)
```

El modelo de Ridge ha demostrado ser altamente preciso y eficaz, lo que lo convierte en una excelente opción para predecir la demanda de personal en los establecimientos industriales en Tierra del Fuego.

Métricas de Evaluación del Modelo

Para el **Modelo de Ridge con Hiperparámetros Ajustados**, presentaremos varias métricas de evaluación para medir su rendimiento. Estas métricas nos ayudarán a entender mejor cómo de bien el modelo está funcionando y a identificar posibles áreas de mejora.

Métricas Utilizadas:

- **Mean Squared Error (MSE):**
 - **Descripción:** El error cuadrático medio es la media de los cuadrados de los errores, es decir, la diferencia entre los valores predichos y los valores reales. Un valor más bajo indica un mejor ajuste del modelo.
 - **Resultado:** 9.108228739648518e-08
- **Mean Absolute Error (MAE):**
 - **Descripción:** El error absoluto medio es la media de las diferencias absolutas entre las predicciones y los valores reales. También un valor más bajo indica un mejor rendimiento del modelo.
 - **Resultado:** 0.00023938539023037582
- **R2 Score:**
 - **Descripción:** El R2 Score, también conocido como coeficiente de determinación, es una medida que indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Un valor más cercano a 1 indica un mejor ajuste.
 - **Resultado:** 0.9999999999997289
- **Mean Absolute Percentage Error (MAPE):**
 - **Descripción:** El error porcentual absoluto medio expresa la precisión como un porcentaje. Proporciona una interpretación fácil de entender de la precisión del modelo.
 - **Resultado:** (Este resultado se puede calcular de la siguiente manera):

```
mape = np.mean(np.abs((y_test - y_pred) / y_test)) * 100
```

```
print("Mean Absolute Percentage Error:", mape)
```

- **Root Mean Squared Error (RMSE):**
 - **Descripción:** La raíz del error cuadrático medio es la raíz cuadrada del MSE. Es útil para interpretar la magnitud del error en la misma escala que los valores originales.
 - **Resultado:** (Este resultado se puede calcular de la siguiente manera):

```
rmse = np.sqrt(mse)

print("Root Mean Squared Error:", rmse)
```

Interpretación de los Resultados:

- **Precision del Modelo:** El Modelo de Ridge con Hiperparámetros Ajustados ha mostrado una precisión excepcional con un R^2 prácticamente perfecto, lo que significa que el modelo puede explicar casi toda la variabilidad de la variable objetivo.
- **Errores Bajos:** Tanto el MSE, MAE, como el RMSE son extremadamente bajos, indicando que las predicciones del modelo están muy cerca de los valores reales.
- **Generalización:** El uso de Ridge Regression ayuda a prevenir el sobreajuste al incluir un término de regularización, lo cual se refleja en la robustez de las predicciones del modelo.