

27/10/2025

Quintana Zepeda Andrea Gpe

Extracción de conocimientos de BD

9C

Reporte: Aplicación de Predicción de Salarios Lab3

1. Objetivo

El objetivo de esta aplicación es permitir a los usuarios:

1. Subir su **notebook (.ipynb)**.
2. Subir un **dataset (.csv)** con los datos de experiencia laboral (YearsExperience) y salarios (Salary).
3. Cargar un **modelo entrenado (.pkl)** de regresión lineal para predecir salarios según los años de experiencia.
4. Visualizar los datos y la predicción en una **tabla y gráfica interactiva**.

2. Herramientas y Tecnologías Utilizadas

- **Python:** Lenguaje principal para el desarrollo.
- **Streamlit:** Framework para crear la interfaz web interactiva.
- **Pandas:** Para la carga, manipulación y visualización de datos tabulares.
- **Matplotlib:** Para crear gráficos de dispersión (scatter plots).
- **Joblib:** Para cargar el modelo de regresión entrenado (.pkl).

3. Estructura de la Aplicación

3.1 Configuración de la página

Se configuró el título, el layout y el encabezado de la página para que sea más visual:

```
st.markdown("""  
<div style="background-color:#8A2BE2;padding:10px;border-radius:10px">  
<h1 style="color:white;text-align:center;">Predicción de Salarios según Años de Experiencia</h1>  
</div>  
""", unsafe_allow_html=True)
```

- Notebook (.ipynb)
- Dataset (.csv)

27/10/2025

Quintana Zepeda Andrea Gpe

Extracción de conocimientos de BD

9C

- Modelo entrenado (.pkl)

```
uploaded_notebook = st.file_uploader("Sube tu Notebook (.ipynb)", type=["ipynb"])
uploaded_csv = st.file_uploader("Sube tu dataset (CSV)", type=["csv"])
uploaded_model = st.file_uploader("Sube tu modelo entrenado (.pkl)", type=["pkl"])
```

3.3 Visualización del dataset

Si el usuario sube un CSV, se muestra:

1. **Tabla de datos** con `st.dataframe(df)`.
2. **Gráfica de dispersión (scatter plot)** de YearsExperience vs Salary.

```
try:
    st.markdown("### Gráfica: YearsExperience vs Salary")
    fig, ax = plt.subplots(figsize=(5,3))
    ax.scatter(df['YearsExperience'], df['Salary'], color="blue", alpha=0.7)
    ax.set_xlabel("YearsExperience")
    ax.set_ylabel("Salary")
    ax.set_title("Relación entre experiencia y salario")
    st.pyplot(fig)
except Exception as e:
    st.error(f"No se pudo graficar: {e}")
```

Se eligió un **tamaño de gráfica reducido** (`figsize=(5,3)`) para que no ocupara demasiado espacio.

3.4 Carga del modelo y predicciones

Se permite al usuario cargar un modelo de regresión lineal guardado en .pkl mediante **joblib**:

```
model = joblib.load(uploaded_model)
```

Para la predicción:

- Se ingresa el número de años de experiencia mediante `st.number_input`.

27/10/2025

Quintana Zepeda Andrea Gpe

Extracción de conocimientos de BD

9C

- Solo se permiten **enteros consecutivos** (step=1).
- Se calcula la predicción con

```
st.markdown("### Predicción de Salario")
experiencia = st.number_input(
    "Años de experiencia",
    min_value=int(df['YearsExperience'].min()) if uploaded_csv else 0,
    max_value=int(df['YearsExperience'].max()) if uploaded_csv else 40,
    step=1
)

if st.button("Predecir salario"):
    try:
        pred = model.predict([[experiencia]])[0]
        st.success(f"Salario predicho: ${pred:,.2f}")
```

3.5 Tabla con predicción

Se genera una nueva fila con la predicción y se concatena al dataset original para mostrar una **tabla completa**:

```
# Tabla con predicción
nueva_fila = pd.DataFrame({
    "YearsExperience": [experiencia],
    "Salary": [pred]
})
if uploaded_csv:
    df_pred = pd.concat([df, nueva_fila], ignore_index=True)
else:
    df_pred = nueva_fila
st.markdown("### Tabla con salarios incluyendo predicción")
st.dataframe(df_pred)
```

27/10/2025

Extracción de conocimientos de BD

Quintana Zepeda Andrea Gpe
9C

4. Características de la Aplicación

1. Gráfica clara
2. Permite predecir salarios solo con años de experiencia enteros.
3. Combina **tabla de datos originales y predicción**.
4. Todo está integrado en una página **interactiva y responsive** usando Streamlit.



5. Conclusión

Se creó una aplicación web funcional para predecir salarios según años de experiencia, combinando **visualización de datos, predicción con modelo de regresión** y manejo de archivos (CSV y Notebook).

El diseño es sencillo, profesional y permite que cualquier usuario pueda:

- Cargar sus datos.
- Visualizar la relación entre experiencia y salario.
- Hacer predicciones rápidas y verlas reflejadas en la tabla de datos.