**Propuesta Detallada del Proyecto**

**1. Objetivo:**

* Desarrollar un modelo predictivo para clasificar el estado oligomérico de las proteínas basado en características estructurales y experimentales.
* Crear un dashboard interactivo que permita explorar proteínas y los resultados del modelo.

**2. Componentes del Proyecto**

1. **Preprocesamiento y Análisis Exploratorio:**
   * Normalizar datos relevantes (ej. Matthews Coefficient, Ligand MW).
   * Codificar variables categóricas (ej. Experimental Method, Crystallization Method).
   * Explorar correlaciones entre características para optimizar el modelo.
2. **Modelo Predictivo:**
   * Algoritmo principal: Random Forest o Gradient Boosting (como XGBoost) para interpretar relaciones complejas.
   * Métricas clave: F1-score, precisión y recall por clase.
3. **Dashboard Interactivo:**
   * Herramientas: **Plotly Dash** o **Tableau** para visualización interactiva.
   * Funcionalidades:
     + Exploración de características de proteínas.
     + Visualización de correlaciones entre variables.
     + Predicciones del modelo, con probabilidad por clase.
     + Filtros dinámicos para proteínas según método de cristalización, resolución, etc.

**3. ¿Cómo Destacar?**

1. **Integración de Métodos Experimentales:**
   * Muchos estudios ignoran datos experimentales como Crystallization Method, lo que hace que este proyecto sea único.
2. **Conexión Visual con Biología:**
   * Los dashboards interactivos son herramientas potentes, pero rara vez se ven en proyectos relacionados con biología estructural.

**4. Implementación del Dashboard**

1. **Carga de Datos:**
   * Importar las predicciones del modelo y las características del dataset original.
2. **Visualizaciones Clave:**
   * Distribución del estado oligomérico por características como Ligand MW o Percent Solvent Content.
   * Diagramas de dispersión, histogramas y gráficos 3D interactivos.
3. **Filtros Dinámicos:**
   * Permitir a los usuarios filtrar proteínas según resolución, método experimental o características del ligando.

**Conclusión**

Este enfoque no solo predice el estado oligomérico, sino que lo convierte en una herramienta accesible y práctica para investigadores. La combinación de machine learning y visualización interactiva eleva la utilidad del proyecto, convirtiéndolo en un ejemplo destacado para un portafolios de analista de datos con enfoque en bioinformática.