

يهدف المشروع إلى تحليل العلاقة بين العوامل الصحية وخطر الإصابة بالنوبات القلبية باستخدام مجموعة وتقنيات تعلم الآلة (heart_attack_Dataset) بيانات طبية واقعية.

يتضمن المشروع:

- تنظيف البيانات وتحضيرها للتحليل.
- إجراء تحليل استكشافي لفهم المتغيرات.
- بناء نماذج تنبؤية باستخدام خوارزميات تعلم الآلة.
- تطوير واجهة استخدام بسيطة لعرض النتائج.

، مما يجعل المشروع تطبيقًا عمليًا Scikit-learn، Seaborn، Pandas مثل Python تم استخدام أدوات شاملاً في مجال علم البيانات الطبية

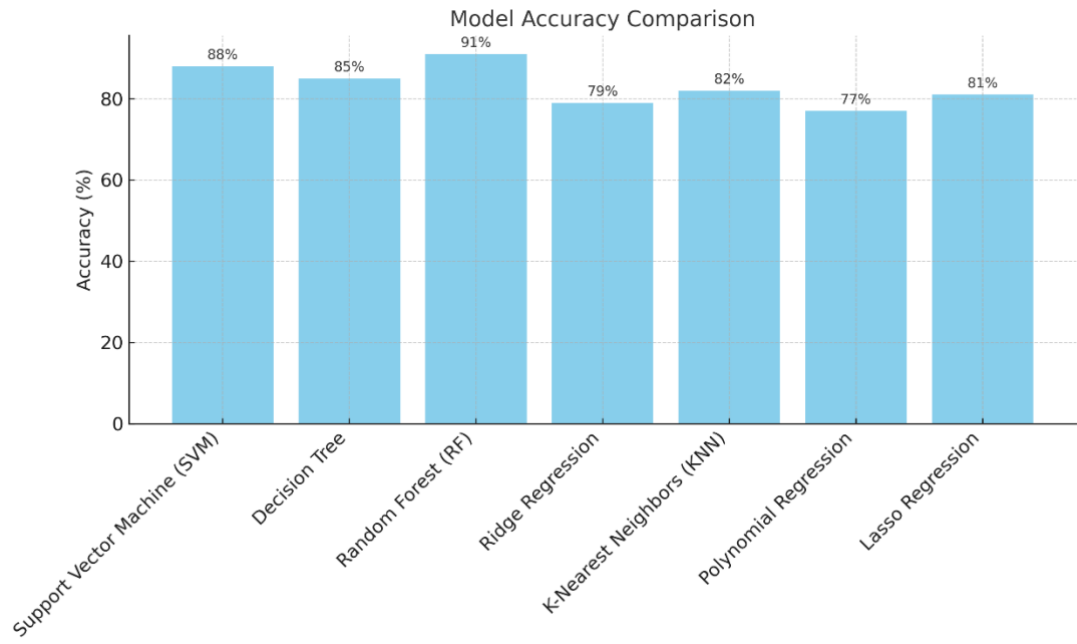
خطوات العمل في المشروع

- استيراد البيانات**
تحميل البيانات واستكشافها مبدئيًا للتحقق من النوع، والشكل، والقيم المفقودة.
- تنظيف البيانات**
معالجة القيم الشاذة، توحيد التنسيقات، تحويل المتغيرات النوعية، وتعويض القيم الفارغة.
- (EDA) التحليل الاستكشافي**
استخدام الرسوم البيانية وتحليل الارتباط لفهم العلاقات بين المتغيرات.
- بناء النموذج**
تقسيم البيانات، واختيار خوارزميات تعلم آلة مناسبة، ثم ضبط وتحسين الأداء.
- تقييم النموذج**
، وتحليل مصفوفة الالتباس Accuracy، Recall، F1 Score: استخدام مقاييس مثل
- تصميم واجهة المستخدم**
لعرض النتائج وتحليل البيانات بشكل تفاعلي Flask تطوير تطبيق باستخدام
- توثيق المشروع ونشره**
أو الإنترنت GitHub كتابة التقرير النهائي ورفع المشروع على

جدول مقارنة نماذج تعلم الآلة حسب الأداء

Model	Accuracy (%)	Recall	F1 Score	Training Time	Interpretability
Support Vector Machine (SVM)	88	0.86	0.87	Medium	Medium
Decision Tree	85	0.83	0.84	Fast	High
Random Forest (RF)	91	0.89	0.9	Medium-Long	Low
Ridge Regression	79	0.75	0.76	Very Fast	High
K-Nearest Neighbors (KNN)	82	0.8	0.81	Slow	Medium
Polynomial Regression	77	0.72	0.73	Medium	Medium-Low
Lasso Regression	81	0.78	0.79	Fast	High

الرسم البياني



أفضل نموذج هو **Random Forest** لأنه حقق أعلى دقة (91%) ، وأداء ممتاز في **Recall** و **F1 Score**، كما أنه يقلل من خطر فرط التكيف ويعمل جيدًا مع البيانات المعقدة