

Pour entraîner notre modèle à prédire l'impact d'un commit sur le code et la probabilité de régression, voici quelques modèles que vous pourriez considérer :

1. Modèles de Classification :

- **Régression logistique** : Bon pour commencer avec des données binaires comme votre classification de régression.
- **SVM (Support Vector Machine)** : Utile pour la classification binaire et multi-classe.
- **Random Forest** : Capable de gérer des données complexes avec interactions non linéaires.
- **Gradient Boosting Machines (GBM)** : Très efficace pour la prédiction et peut gérer de grandes quantités de données avec des relations complexes.
- **Réseaux de neurones artificiels (ANN)** : Pour les modèles plus complexes nécessitant une compréhension plus profonde des données.

2. Modèles de Régression (pour prédire l'impact quantitatif) :

- **Régression linéaire** : Pour modéliser la relation entre les variables explicatives et la réponse quantitative.
- **Régression Ridge et Lasso** : Pour gérer la multicollinéarité et la régularisation.
- **Réseaux de neurones profonds (DNN)** : Pour des prédictions très complexes et non linéaires basées sur des données riches en fonctionnalités.

3. Modèles de Séquence Temporelle (si la chronologie des commits est cruciale) :

- **LSTM (Long Short-Term Memory)** : Pour modéliser les dépendances temporelles et les séquences.
- **GRU (Gated Recurrent Unit)** : Une alternative plus légère aux LSTM pour le traitement des séquences.

4. Modèles combinés ou ensemblistes :

- **Modèles de stacking** : Combinaison de plusieurs modèles pour améliorer la précision prédictive.
- **Modèles de bagging et de boosting** : Pour agréger les prédictions de plusieurs modèles de base.

5. Modèles d'Interprétabilité :

- **Arbres de décision** : Pour comprendre les décisions de modèle et les relations entre les variables.
- **SHAP (SHapley Additive exPlanations)** : Pour expliquer les prédictions des modèles de manière globale ou individuelle.