

Model planning

Predict whether a motion to strike in Connecticut civil cases is granted or denied.

- Court Administrative Data (from CT Judicial Branch):
 - Attorney ID (juris number)
 - Attorney specialization (computed using entropy of case types)
 - Case type (tort vs. vehicular)
 - Case location (which court)
- Goal: Predict whether a motion is *granted* or *denied* (your MotionResultCode column).
- Type: This is a binary classification problem.

metadata features (no text yet). From your CSV, start with:

- Case info: CaseLocation, CaseMajorCode, CaseMinorCode
- Motion info: MotionJurisNumber, MotionDocumentTypeName
- Attorney/judge info: CaseAttorneyType, CaseDispositionJudgeJurisNo
- Outcome (target): MotionResultCode
- motion to strike is successful or denied
- We are trying to find trends to see what indicators help show that theres a motion to strike or deny

	Actual positive	Actual negative
Predicted positive	True positive (TP): Motion to Strike successful, is true	False positive (FP) Says Motion to Strike successful, is actually denied
Predicted negative	False negative (FN): Says Motion to Strike is denied, actually passed	True negative (TN): Says Motion to Strike is denied, is actually denied

- We want to Maximize the TP and TN as much as we can, I'd say we would rather have a False Negative than a false positive? but FN and FP are still about the same weight in terms of risk

What Model will do:

1. Train model with dataset, given test
2. Make predictions based on each case,
3. Compare with the actual outcomes, Determine the TP, FP, FN, TN,
4. Determine the Accuracy of the model, can also display the TPR, FPR, Precision for data sake