**Final Project Proposal: Classifying Retirement Plan Financial Adequacy Using Form 5500 Data**

**Context & Motivation**

Employer-sponsored retirement plans, especially 401(k)s, are the primary savings vehicle for millions of Americans. But not all plans equally support long-term financial security. Some offer strong structural support for participant savings, while others suffer from low engagement, high leakage via early withdrawals, or poor contribution patterns.

The U.S. Department of Labor collects detailed annual filings from these plans via Form 5500, which include financial, demographic, and operational data. These filings offer a unique opportunity to assess whether a retirement plan is structurally enabling participants to save adequately for retirement and whether macroeconomic conditions compromise that readiness.

**Dataset Selection**

I propose to use publicly available Form 5500 filings from the Department of Labor’s EFAST2 portal, covering the years 2020–2024. Each filing contains structured data on plan assets, contributions, distributions, participant counts, administrative expenses, and more. I will filter the dataset to focus final (complete) filings for defined contribution plans (primarily 401(k)s), which are the most common retirement vehicle in the private sector.

**Learning Question**

Can we classify retirement plans as “adequate” or “inadequate” in supporting participant financial readiness, based on structural indicators such as participation rate, average account balance, contribution stability, and leakage volume?

This question is relevant to employers that sponsor these plans, policymakers, and retirement service providers who want to understand whether a plan is structurally supporting participant financial security, especially in the face of economic disruptions like inflation, unemployment, or market volatility.

**Dataset Details**

The dataset includes approximately 240,000 plan filings across five years. Each row represents a plan-year filing, and I will use a combination of Employer Identification Number (EIN) and Plan Number to track continuity across years.

This is a classification problem, with the target variable defined as a binary label: adequate vs. inadequate, based on engineered thresholds for savings adequacy, engagement, and leakage.

The data, which can be viewed via the following link, is divided into four datasets, one for the Form 5500 filing and three supporting datasets for Schedules H, I and R.

To classify retirement plans as structurally adequate or insufficient, I plan to use the 5500 data to engineer features across six diagnostic categories: participant engagement, savings behavior, leakage and access, plan structure, temporal sensitivity, and fairness overlays. These features will capture how well a plan enables consistent saving, accommodates financial stress, and supports participant inertia. Ratios like participation rate, contribution per participant, and leakage burden offer interpretable signals of plan adequacy, while stratified diagnostics across industries and plan sizes support fairness and stakeholder relevance.

**Model Selection**

To build a principled and interpretable classifier, I plan to explore:

**Logistic Regression:** For baseline interpretability and fairness audits.

**Random Forest:** To capture nonlinear relationships and assess feature importance.

Each model will be evaluated using ROC/AUC, confusion matrices, and fairness overlays across industries and plan sizes. I’ll also use residual diagnostics to identify misclassified plans and explore macroeconomic sensitivity.

**GitHub Repository**

I’ll be sharing my data, code, documentation, and findings here: [GitHub Repo Link] — (link will be added once the repository is live)

**Data Leakage Avoidance**

This project avoids data leakage by using only plan-level features available at the time of filing. No future outcomes, external economic indicators, or derived labels from post-filing behavior are used in training. All engineered features, e.g. participation rate, average account balance, leakage ratio, are constructed from within-year data. Temporal features across years are used cautiously and only when modeling longitudinal trends.

**Question Significance**

The learning question, “Can we classify retirement plans as structurally adequate or insufficient in supporting participant financial readiness?” is highly impactful. It addresses a real-world challenge faced by sponsors, policymakers, and service providers: identifying which plans are enabling long-term financial security and which may be structurally failing participants. The insights can inform outreach, redesign, and policy interventions.

**Potential for Misuse**

There is a risk that the classifier could be misinterpreted as a judgment of individual participant behavior or investment performance. To mitigate this, the model is explicitly framed as a structural diagnostic tool, not a predictor of individual outcomes. Misuse could also arise if sponsors use the model to justify disengagement from vulnerable plans rather than improving them. Clear documentation and ethical framing are essential.

**Investment Data**

While investment selection and performance are undeniably a component of retirement readiness, this analysis intentionally excludes direct investment-level data, e.g. asset allocation, fund returns, volatility, from the classification model. The rationale is twofold:

**Structural Focus:** This project emphasizes plan-level structural indicators, such as participation, contributions, leakage, and fee burden, that reflect how well a plan enables participants to save. These are actionable, interpretable, and directly tied to sponsor decisions.

**Data Limitations:** Investment details in Form 5500 filings (e.g., Schedule H line items) are often inconsistently reported, difficult to normalize across plans, and not reliably attributable to participant outcomes without individual-level data.

By focusing on structural adequacy rather than investment performance, the model aims to provide a principled, diagnostic view of retirement plan readiness that is both reproducible and stakeholder-relevant.

**Lack of Participant-Level Financial Context**

One potential weakness is that this model cannot account for individual participant financial circumstances, such as income, debt, employment status, or household dynamics, that directly influence contribution and withdrawal behavior. While the classifier uses plan-level structural indicators, e.g. participation rate, average balance, leakage volume, it cannot infer whether low savings or high leakage is due to plan design or external financial hardship.

This limitation is intentional: Form 5500 data does not include individual-level financial information, and the model is designed to assess plan adequacy, not participant intent. However, this means the classifier may occasionally flag structurally sound plans as “insufficient” if participant behavior reflects broader economic stress rather than plan design flaws.

**Behavioral and Structural Context**

Another concern is that participation in retirement plans is shaped not only by plan design but by the realities of financial wellness and behavioral inertia.

For many individuals, the ability to access funds through loans or hardship withdrawals is essential. Locking up assets in a retirement plan without this flexibility can deter enrollment, especially in financially vulnerable populations.

Moreover, inertia plays a powerful role: if a participant is not automatically enrolled and contributions are not deducted from payroll, saving requires deliberate action. Most people don’t take that step.

These behavioral dynamics mean that low participation or contribution rates may reflect financial constraints or plan accessibility issues, not simply a lack of retirement readiness. This underscores the importance of interpreting model outputs as structural diagnostics, not judgments of individual behavior.

**COVID Impact**

The COVID-19 pandemic introduced significant disruptions to retirement plan operations, and these effects are reflected in Form 5500 filings from 2020 and 2021. Regulatory relief allowed plan sponsors to delay forwarding participant contributions and loan repayments without triggering enforcement violations, provided the delays were pandemic-related. As a result, some filings may underreport contribution timing issues or omit delinquency disclosures that would otherwise appear on Schedule H.

Participant counts were also affected. Furloughed employees created ambiguity in reporting: depending on plan design and sponsor interpretation, some were counted as active participants while others were excluded or reported under COBRA continuation coverage. This variability complicates the interpretation of engagement metrics and may distort participation-based adequacy labels.

For modeling purposes, these years require careful handling. Plans may appear structurally “inadequate” due to temporary disruptions rather than enduring design flaws. To account for this, COVID-era filings should be flagged and considered separately in fairness audits, residual diagnostics, and temporal drift analyses. This ensures that the classifier remains ethically grounded and interprets adequacy in context, and not in isolation from the economic realities of the time.

**Ethical Considerations**

This project will emphasize fairness, transparency, and stakeholder relevance. Classification will be done at the plan level to avoid privacy concerns. The model will avoid attributing readiness to uncontrollable market forces and instead focuses on actionable plan traits. Ethical modeling choices will be documented throughout.