User Needs and Defining Success

Corporate Financial Distress Early-Warning System

Project Group 22:

1. Bryan Zeng
2. Harshit Raheja
3. Sankalp
4. Piriyajeishree Murali Naidu
5. Nandana Pradeep
6. Palak Tanwar

1. Evidence of User Need

User Research Summary

Date	Source	Summary of Findings
2024	Academic Literature (Altman Z-Score Studies)	Traditional linear models (Altman Z-Score from 1968) fail to capture complex, non-linear relationships in modern financial data. Static thresholds don't adapt to changing macroeconomic conditions.
2024	Credit Rating Agency Analysis	Credit ratings from Moody's and S&P are slow to update and "sticky." Ratings often don't reflect rapid deterioration until official downgrade announcements, leaving stakeholders with delayed signals.
2024	SEC EDGAR Filing Analysis	Financial distress signals exist in 10-K/10-Q filings months before official bankruptcy. Warning signs are dispersed across multiple reports and difficult to interpret manually at scale (10,000-20,000 companies).
2024	Industry Practice Review	Financial analysts cannot manually monitor thousands of companies continuously. Coverage gap exists where only large-cap companies are scrutinized; small-to-mid-cap companies often fail without early warning.
2024	UCLA LoPucki Bankruptcy Database	Historical bankruptcy data shows distress rate of 2-5% among public companies. Pattern analysis reveals detectable financial deterioration patterns 6-12 months before official events.

How might we solve:

Early detection and prediction of corporate financial distress (bankruptcy, liquidity crises, or involuntary delisting) 6-12 months in advance, using publicly available financial and macroeconomic data, to provide stakeholders with actionable, interpretable risk signals before severe financial deterioration becomes public.

AI Probably Better	AI Probably NOT Better
The core experience requires prediction of future events. → Predicting distress 6-12 months ahead Need to detect low occurrence events that are constantly evolving. → Distress events are rare (2-5%) and patterns shift with economic conditions Need to recognize a general class of things too large to articulate every case. → Complex non-linear interactions between 40-80 financial features The user experience doesn't rely on predictability. → Users expect probabilistic risk scores, not deterministic answers Personalization will improve the user experience.	The most valuable part is its predictability regardless of context. → Context (macroeconomic conditions) IS critical The cost of errors is very high and outweighs benefits. → System is decision-support, not automated action Users need to understand exactly everything in the code. → SHAP explanations provide interpretability People explicitly tell you they don't want automation. → Stakeholders want scalable monitoring

We think AI CAN help solve early detection of corporate financial distress, because:

1. The problem fundamentally requires prediction of future events (distress 6-12 months ahead), which is a core AI strength.

2. Distress events are rare (2-5%) and evolving, requiring pattern recognition across complex, non-linear feature interactions that static rules cannot capture.

3. Scale is impossible for humans: monitoring 10,000-20,000 companies continuously with 40-80 features each exceeds manual analysis capacity.

4. The system augments rather than replaces human judgment, providing interpretable risk scores with SHAP explanations as decision-support.

2. Augmentation versus Automation

Based on stakeholder analysis, we identified the following user attitudes for our target users:

User Type	Attitude Toward Automation	Preference
Investors	Want early warnings but need to understand WHY before acting on recommendations	Augmentation - Risk scores with explanations

Lenders/Creditors	Regulatory requirements demand human oversight; cannot auto-reject based solely on ML predictions	Augmentation - Decision support with audit trail
Regulators	Need comprehensive market monitoring but require interpretable evidence for interventions	Augmentation - Systemic risk dashboards
Financial Analysts	Want automated screening to reduce manual workload but retain final judgment authority	Augmentation - Prioritized watchlists

Research Protocol Questions (Adapted for Our Context)

- If you were training a new analyst, what would be the most important tasks to teach first?

  Response: Reviewing financial ratios, understanding trend analysis, recognizing red flags in cash flow
- How often do you review company financial health?

  Response: Quarterly (after 10-Q/10-K releases) - too infrequent for early warning
- If you had an AI assistant for this task, what duties would you give it?

  Response: Continuous monitoring, flagging anomalies, prioritizing which companies need attention

Conclusion: All user segments prefer AUGMENTATION over full automation. The system should:

- Provide risk scores and alerts as decision-support, not automated actions
- Include SHAP-based explanations for every prediction
- Position clear disclaimers that this is an educational/analytical tool, not financial advice

3. Design Your Reward Function

Weighing the trade-offs between precision and recall for the user experience:

	Prediction: POSITIVE (Flagged as High Risk)	Prediction: NEGATIVE (Not Flagged)
Reference: POSITIVE (Actually Distressed)	TRUE POSITIVE Company correctly flagged before bankruptcy Example 1: Flagged company files bankruptcy 8 months later Example 2: Flagged company gets delisted 6 months later Example 3: Flagged company experiences severe liquidity crisis	FALSE NEGATIVE Missed a company that actually failed Example 1: Company goes bankrupt with no warning Example 2: Sudden delisting surprises investors Example 3: Creditors lose money on unexpected default
Reference: NEGATIVE (Actually Healthy)	FALSE POSITIVE Healthy company incorrectly flagged Example 1: Flagged company recovers, no distress Example 2: Temporary dip misclassified as risk	TRUE NEGATIVE Healthy company correctly not flagged Example 1: Stable company remains unflagged Example 2: Growing company correctly assessed

	Example 3: Industry-wide downturn triggers false alarms	Example 3: Well-capitalized firm not alarmed

Precision vs Recall Trade-off Analysis

Cost of False Negatives (Missed Distress):

• Investors lose significant capital in unexpected bankruptcies
• Lenders face default losses without time to restructure
• System credibility destroyed if major failures are missed

Cost of False Positives (False Alarms):

• Wasted analyst time investigating healthy companies
• Alert fatigue if too many false alarms
• Potential reputational harm to incorrectly flagged companies (mitigated by decision-support positioning)

Our AI model will be optimized for recall because:

Missing an actual distress event (false negative) has catastrophic consequences: investors lose capital, lenders face unexpected defaults, and the system loses all credibility. The primary value proposition is providing early warning, which is destroyed if distressed companies are missed.

We understand the trade-off:

Our model will generate more false positives (healthy companies flagged as risky). This means analysts will need to investigate more companies that turn out to be fine. We mitigate this through: (1) Precision@K metrics for top-risk companies, (2) SHAP explanations to help analysts quickly validate/dismiss alerts, (3) Tiered alert thresholds (high/medium/low risk) to prioritize investigation.

4. Define Success Criteria

Success Metrics Framework

Version 1: Primary ML Performance Metric

If ROC-AUC for the distress prediction model drops below 0.80 on the test set, we will trigger a retraining pipeline with updated data and hyperparameter tuning.

Version 2: High-Risk Company Precision

If Precision@K (top 100 highest-risk companies) for the alert system drops below 80%, we will review threshold calibration and feature engineering to reduce false positives in the high-priority tier.

Version 3: Model Drift Detection

If feature distribution drift (measured by PSI) for key financial ratios goes above 0.25, we will flag for investigation and schedule evaluation against recent labeled data.

Statement Iteration

- Is this metric meaningful for all of our users?

Yes - ROC-AUC and Precision@K translate to: "Can we trust the alerts?" and "Are we catching real problems?"

- How might this metric negatively impact some of our users?

Optimizing for recall may increase analyst workload investigating false positives. Mitigated by tiered alerts and SHAP explanations.

- Is this what success means for our feature on day 1?

Yes - achieving ROC-AUC ≥ 0.80 validates the core prediction capability.

What about day 1,000?

At scale, we add business metrics: lead time before distress, user trust surveys, and actual loss prevention estimates.

Final Version: Comprehensive Success Criteria

Technical Metrics:

• Target ROC-AUC ≥ 0.80 on test set (2022-2023 data)
• Precision@K ≥ 80% for top-risk companies
• Brier Score < 0.15 for probability calibration

Operational Metrics:

• Automated end-to-end pipeline execution without manual intervention
• API response time < 2 seconds for single-company queries
• Model explanations (SHAP) interpretable to domain experts

Business Metrics (Future):

• 6-12-month lead time between alert and actual distress event
• Alert precision validated against actual outcomes quarterly
• Schedule Regular Reviews

Success Metric Review Schedule

Weekly Reviews (During Development):

• Training metrics convergence
• Validation set performance trends

Phase Milestone Reviews:

• End of Phase 3 (Model Training): Initial ROC-AUC evaluation
• End of Phase 4 (Evaluation): Full metric suite on test set

Post-Deployment (Quarterly):

• Drift monitoring and performance degradation checks
• Comparison of predicted vs. actual distress events