Model Predictie Insolventa: I-ML-v2026.32

(Data: 15.06.2026, Versiune document: 1.0)

I. Introducere

Modelul *I-ML-v2026.32* este un model bazat pe tehnici de *machine learning (ML)* [1], dezvoltat pentru estimarea probabilității de insolvență a unei companii. Modelul utilizează informații financiare istorice și generează estimări probabilistice care pot sprijini procesul decizional. Rezultatele sale trebuie interpretate ca indicatori de risc și nu ca certitudini privind evoluția viitoare a unei firme.

Evaluarea modelului a fost realizată pe date furnizate de ListaFirme (https://listafirme.ro), care includ bilanțuri financiare anuale și informații privind starea de insolvență a companiilor din România. Performanța a fost măsurată folosind metrici standard, precum *acuratețea*, *precizia* și *senzitivitatea* (https://en.wikipedia.org/wiki/Precision_and_recall), pe seturi de date independente de antrenare.

Rezultatele pentru anul 2025 arată că *I-ML-v2026.32* depășește semnificativ performanța modelului Altman Z-Score [2], [3], precum și a unor modele clasice ML. Astfel, modelul identifică mai eficient companiile cu probabilitate de insolvență și transformă datele financiare în avantaj competitiv.

Notă importantă

I-ML-v2026.32 este un model statistic/ML, iar rezultatele sale reprezintă estimări ale probabilității de insolvență, nu certitudini. Acestea trebuie utilizate ca suport decizional, împreună cu alte informații relevante și analiza profesională a utilizatorului.

II. Date & Scenariu Evaluare

Modelul *I-ML-v2026.32* a fost dezvoltat și evaluat utilizând date reale furnizate de ListaFirme (https://listafirme.ro), care includ bilanțuri financiare anuale și informații privind insolvența companiilor din România. Estimarea probabilității de insolvență reprezintă o problemă dificilă atât din perspectivă economică, cât și statistică, deoarece cazurile de insolvență sunt relativ rare în raport cu numărul total de firme active. În plus, pentru majoritatea companiilor românești nu sunt disponibile informații bursiere, ceea ce limitează sursele de informație utilizabile pentru evaluarea riscului financiar.

Pentru a evalua în mod obiectiv performanța modelului *I-ML-v2026.32*, acesta a fost comparat cu o serie de modele economice și ML consacrate, utilizând același set de date de testare și același orizont de predicție de un an.

Setul de testare a fost construit folosind informații aferente bilanțurilor financiare raportate pentru anul 2024, iar starea de insolvență a fost determinată pe baza evenimentelor înregistrate în anul 2025. Acesta conține 947.846 de companii distincte pentru care au fost disponibile toate informațiile necesare analizei. Dintre acestea, doar pentru 4.277 de companii (aprox. 0,5%) a fost deschis un proces de insolvență în anul 2025, evidențiind caracterul puternic dezechilibrat al problemei. Experimentul reproduce un scenariu realist în care modelele au acces doar la informațiile disponibile la momentul efectuării predicției. Astfel, niciun model evaluat nu a avut acces la date financiare sau informații privind insolvențele din anul următor.

Modelul *I-ML-v2026.32* disponibil în prezent pe platforma ListaFirme (https://listafirme.ro/scor-I-ML-v2026.32enta.asp) a fost ulterior reantrenat prin includerea bilanțurilor financiare din 2024 și a informațiilor privind insolvențele observate în anul 2025, beneficiind astfel de un set de date mai recent decât cel utilizat în experimentul de evaluare prezentat în acest document. În plus, atunci când cele mai recente date financiare ale unei companii nu sunt disponibile, este utilizat un model cu performanțe similare, care estimează probabilitatea de insolvență pe un orizont de predicție de doi ani.

III. Modele Benchmark

Performanța modelului *I-ML-v2026.32* a fost comparată cu modele de referință utilizate frecvent în evaluarea riscului de insolvență. Aceste benchmark-uri includ atât modele economice consacrate, cât și algoritmi fundamentali ML.

Modelul Altman. Modelul Altman este unul dintre standardele de referință pentru evaluarea riscului de insolvență al companiilor. Acesta utilizează o combinație de indicatori financiari pentru a estima probabilitatea ca o firmă să intre în insolvență [2], [3]. În această analiză, folosim o versiune adaptată pentru companiile private din România.

Modelul LR. *Logistic Regression (LR)* este unul dintre cele mai utilizate modele statistice pentru estimarea probabilității producerii unui eveniment. Modelul calculează probabilitatea de insolvență pe baza unei combinații de indicatori financiari și reprezintă un reper fundamental în literatura de specialitate și în aplicațiile de ML (https://en.wikipedia.org/wiki/Logistic_regression). În cadrul acestui experiment, am utilizat implementarea LR din biblioteca *scikit-learn* pentru Python (https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html). Antrenarea și evaluarea au fost realizate pe seturile de date descrise anterior.

Notă: Pentru a ține cont de dezechilibrul dintre firmele solvabile și cele aflate în proces de insolvență, modelul a fost configurat cu parametrul **class_weight = balanced**. În rest, au fost utilizate setările implicite ale algoritmului din librăria scikit-learn.

Modelul DT. *Decision Tree (DT)* este un model fundamental ML, care construiește reguli de decizie pe baza exemplelor observate în date. Spre deosebire de modelele liniare, un arbore de decizie nu presupune o formă fixă a relației dintre variabilele de intrare și rezultatul estimat, ceea ce îi permite să surprindă relații mai complexe între indicatorii financiari și probabilitatea de insolvență (https://en.wikipedia.org/wiki/Decision_tree_learning). În cadrul acestui experiment, am utilizat implementarea DT din biblioteca *scikit-learn* pentru Python (https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html). Antrenarea și evaluarea au fost realizate pe seturile de date descrise anterior.

Notă: Pentru a ține cont de dezechilibrul dintre firmele solvabile și cele aflate în proces insolvență, modelul a fost configurat cu parametrul **class_weight = balanced**. Adâncimea maximă a arborelui a fost limitată la **max_depth = 10**, pentru a reduce riscul de supraînvățare și pentru a îmbunătăți capacitatea de generalizare. În rest, au fost utilizate setările implicite ale algoritmului din librăria scikit-learn.

IV. Metrice Evaluare

În evaluarea modelelor, am utilizat mai multe metrici standard, recunoscute în literatura de specialitate. Acestea se bazează, în principal, pe cele patru elemente ale matricei de confuzie, utilizată în evaluarea modelelor de clasificare binară:[1] *true negatives (TN)*, *true positives (TP)*, *false negatives (FN)* și *false positives (FP)* (https://en.wikipedia.org/wiki/Confusion_matrix). În plus față de evaluarea clasificării, am analizat și calitatea probabilităților estimate de modele, comparând probabilitatea de intrare în insolvență cu rezultatele observate în datele istorice. Pentru această evaluare am utilizat metrici standard precum *Brier Score* și *Log Loss* [4], [5].

Notă importantă

Metricile prezentate descriu performanța modelului pe seturile de date utilizate în această analiză. Acestea nu constituie o garanție privind performanța viitoare a modelului în alte contexte sau aplicații. Orice decizie de business trebuie fundamentată și pe alte informații relevante și validări independente realizate de utilizator.

Acuratețea (Acc). Acuratețea măsoară proporția predicțiilor corecte din totalul predicțiilor realizate de model (https://en.wikipedia.org/wiki/Accuracy_and_precision):

$$\mathrm{Acc} = \frac{TP + TN}{TP + TN + FP + FN}.$$

[1]Aici, companii în proces de insolvență (1) și companii solvabile (0).

Deși este una dintre cele mai utilizate metrici de evaluare, acuratețea poate fi înșelătoare în cazul seturilor de date puternic dezechilibrate. De exemplu, dacă 99% dintre companii sunt solvabile, un model care prezice întotdeauna că o companie este solvabilă va obține o acuratețe de 99%, fără a identifica însă niciun caz real de insolvență.

Precizia (Prec). Precizia măsoară proporția companiilor identificate de model ca având probabilitate mare de insolvență care au intrat efectiv în proces de insolvență (https://en.wikipedia.org/wiki/Precision_and_recall):

$$\mathrm{Prec} = \frac{TP}{TP + FP}.$$

O valoare ridicată a preciziei indică un număr redus de alarme false (FP), în timp ce o valoare scăzută sugerează că modelul semnalează frecvent drept riscante companii care nu vor intra în proces de insolvență.

Senzitivitatea (Rec). Senzitivitatea, cunoscută și sub denumirea de *Recall*, măsoară capacitatea modelului de a identifica companiile care au intrat efectiv în proces de insolvență (https://en.wikipedia.org/wiki/Precision_and_recall):

$$\mathrm{Rec} = \frac{TP}{TP + FN}.$$

O valoare ridicată a senzitivității indică faptul că modelul detectează o proporție mare dintre companiile care intră în insolvență. În schimb, o valoare scăzută sugerează că modelul ratează multe cazuri reale de insolvență (FN).

Scorul F1 (F1). Scorul F1 reprezintă media armonică dintre precizie și senzitivitate (https://en.wikipedia.org/wiki/F-score):

$$\mathrm{F1} = 2\frac{Prec \times Rec}{Prec + Rec}.$$

Această metrică este utilă atunci când precizia și senzitivitatea sunt la fel de importante. Scorul F1 este utilizat frecvent în problemele cu clase dezechilibrate, deoarece evaluează simultan capacitatea modelului de a identifica corect cazurile pozitive și de a limita alarmele false, fiind mai puțin influențat de numărul mare de exemple false clasificate corect (TN).[2]

Coeficientul de Corelație Matthews (MCC). *Matthews Correlation Coefficient* este o metrică care ia în considerare toate elementele matricei de confuzie și oferă o evaluare echilibrată a performanței modelului, inclusiv în cazul seturilor de date dezechilibrate (https://en.wikipedia.org/wiki/Phi_coefficient):

$$\mathrm{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}.$$

În general, MCC măsoară gradul de concordanță dintre predicțiile modelului și rezultatele reale. O valoare ridicată indică o capacitate bună de discriminare între companiile care intră în insolvență și cele care rămân solvabile, în timp ce o valoare apropiată de zero sugerează performanțe comparabile cu o clasificare aleatorie.

[2]Aici, numarul de firme solvabile clasificate corect.

Scorul Brier (BrierScore). Scorul Brier măsoară calitatea probabilităților estimate de un model, cuantificând diferența dintre rezultatele observate și probabilitățile prezise pentru întregul set de date (https://en.wikipedia.org/wiki/Brier_score):

$$\text{BrierScore} = \frac{1}{N}\sum_{i=1}^{N}(y_i - p_i)^2,$$

unde N reprezintă numărul de observații, y_i este clasa reală, iar p_i este probabilitatea estimată de model pentru observația i. O valoare mai mică a scorului Brier indică probabilități mai bine calibrate și, implicit, o calitate mai bună a estimărilor furnizate de model.

Eroarea Logaritmică (LogLoss). *Log Loss* măsoară calitatea probabilităților estimate de un model, penalizând mai sever predicțiile incorecte realizate cu un grad ridicat de încredere (https://en.wikipedia.org/wiki/Cross-entropy):

$$\text{LogLoss} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\right],$$

unde N reprezintă numărul de observații, y_i este clasa reală, iar p_i este probabilitatea estimată de model pentru observația i. O valoare mai mică a metricii *LogLoss* indică probabilități mai bine calibrate și, implicit, o performanță mai bună a modelului în estimarea probabilitatii de insolvență.

V. Rezultate

În simularea aferentă anului 2025, *I-ML-v2026.32* a obținut cele mai bune rezultate generale dintre toate modelele evaluate, depășind atât modelul Altman, cât și modele ML fundamentale.

Tabelele de mai jos prezintă rezultatele obținute de modelele evaluate atât din perspectiva clasificării companiilor cu probabilitate mare de insolvență, cât și din perspectiva calității probabilităților estimate. Având în vedere dezechilibrul puternic al datelor, metrici compuse precum F1 și MCC oferă o imagine mai relevantă asupra performanței reale decât simpla acuratețe. Modelul *I-ML-v2026.32* obține cel mai bun echilibru între identificarea companiilor cu probabiliate mare de insolvență și limitarea alarmelor false, si furnizează cele mai precise estimări probabilistice.

TABEL I
PERFORMANȚA CLASIFICĂRII

Model	Acc	Prec	Rec	F1	MCC
Predicție Random	**0.995**	0.005	0.005	0.005	0.000
Altman	0.990	**0.293**	0.110	0.160	0.178
LR	0.815	0.016	0.664	0.031	0.082
DT	0.955	0.077	**0.822**	0.141	0.243
I-ML-v2026.32	0.993	0.288	0.616	**0.393**	**0.418**

Notă: *Datorită numărului redus de companii aflate în proces de insolvență, acuratețea nu este o metrică relevantă pentru compararea performanței modelelor. Metricile compuse F1 și MCC oferă o imagine mai fidelă a calității predicțiilor.*

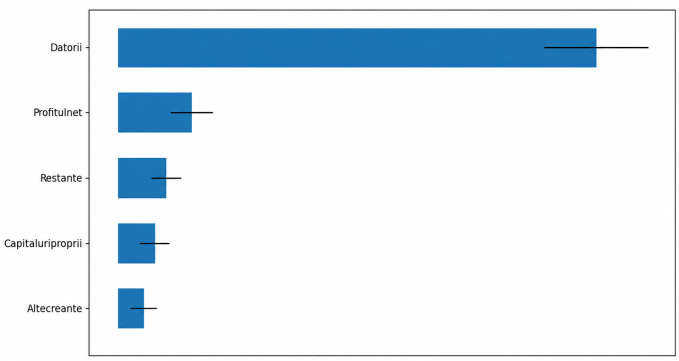
TABEL II
CALITATEA ESTIMĂRILOR PROBABILISTICE

Model	BrierScore	LogLoss
Predicție Random	0.004	0.029
Altman	0.247	0.688
LR	0.145	0.470
DT	0.040	0.156
I-ML-v2026.32	**0.004**	**0.015**

VI. Concluzie & Explicații avansate

Modelul *I-ML-v2026.32* utilizează tehnici avansate ML pentru a identifica tipare și relații în datele financiare ale companiilor. Rezultatele prezentate în acest document arată că modelul oferă o capacitate superioară de estimare a probabilității de insolvență comparativ cu modelele de referință evaluate pe datele furnizate de ListaFirme (https://listafirme.ro).

Deși modelul analizează simultan un număr mare de indicatori financiari, anumite elemente din bilanț au o influență mai mare asupra estimării finale a probabilității de insolventa. Înțelegerea acestor factori poate oferi informații suplimentare despre punctele forte și vulnerabilitățile financiare ale unei companii. Figura 1 prezintă indicatori importanți de bilanț utilizați de model în estimarea probabilității de insolvență.



Imagine 1. Importanța datelor de bilanț *(I-ML-v2026.32)*

Notă importantă

Modelul *I-ML-v2026.32* are o natură statistică, bazată pe tehnici de *machine learning (ML)* [1]. În consecință, rezultatele generate pe baza datelor disponibile reprezintă estimări probabilistice și nu constituie certitudini privind evoluția viitoare a unei companii. Modelul este destinat exclusiv ca instrument de suport decizional și nu trebuie utilizat ca unic temei pentru acordarea de credite, realizarea de investiții sau adoptarea altor decizii cu impact juridic ori financiar semnificativ. Orice decizie trebuie fundamentată și pe alte informații relevante, precum și pe analiza și judecata profesională a persoanelor responsabile.

SURSE

[1] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[2] E. I. Altman, "Predicting financial distress of companies: revisiting the z-score and zeta® models," in *Handbook of research methods and applications in empirical finance*. Edward Elgar Publishing, 2013, pp. 428–456.

[3] D. D. L. GRIGORESCU, "Model de analiză discriminantă pentru detectarea riscului de faliment," *Revista Română de Statistică-Supliment nr*, p. 33, 2019.

[4] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 625–632.

[5] T. Silva Filho, H. Song, M. Perello-Nieto, R. Santos-Rodriguez, M. Kull, and P. Flach, "Classifier calibration: a survey on how to assess and improve predicted class probabilities," *Machine Learning*, vol. 112, no. 9, pp. 3211–3260, 2023.