

ADABOOST-ALAPÚ MEGKÖZELÍTÉS MOTORHIBÁK ELŐREJELZÉSÉRE

Hornyák Olivér 

egyetemi docens, Miskolci Egyetem, Informatikai Intézet, Alkalmazott Informatikai Intézeti Tanszék
3515 Miskolc-Egyetemváros, e-mail: oliver.hornyak@uni-miskolc.hu

Absztrakt

A motorok meghibásodásának pontos előrejelzése különösen kritikus olyan iparágakban, ahol az eszközök kiesése jelentős költségeket és biztonsági kockázatokat eredményezhet. Ez a tanulmány az adaptív boosting (AdaBoost) algoritmus alkalmazását vizsgálja a motorok meghibásodásának előrejelzésére. Az AdaBoost választását az is indokolta, hogy könnyen értelmezhető és kis számítási erőforrásigényű, így jól alkalmazható valós idejű vagy beágyazott rendszerekben történő implementációra. A cikkben vizsgált adathalmaz 100 motor működési adatait tartalmazza. Az előfeldolgozás során a motorok meghibásodási idejének (Time to Failure, TTF) mediánja alapján bináris címkéket hoz létre, amely jelzi a karbantartási igény sürgősségét. Az implementált modell 80%-os pontosságot ért el. Az eredmények alapján a modell előrejelzési képessége ígéretes, de a hamis negatív osztályzások száma még további optimalizálást igényel. A kutatás eredményei új lehetőségeket nyitnak a megelőző karbantartási rendszerek fejlesztésére.

Kulcsszavak: AdaBoost, prediktív karbantartás, mesterséges intelligencia, meghibásodás-előrejelzés, TTF

Abstract

This study investigates the application of the adaptive boosting (AdaBoost) algorithm for predicting engine failures, a critical task in industries where equipment downtime incurs significant costs and safety risks. The choice of AdaBoost was motivated by its interpretability and low computational requirements, making it well-suited for real-time or embedded system implementations. The analysis utilized a dataset comprising operational data from 100 engines. During preprocessing, binary labels were created based on the median of the engines' Time to Failure (TTF) to indicate the urgency of maintenance needs. The implemented model achieved an overall accuracy of 80%. The results demonstrate the model's promising predictive capabilities, though the number of false negatives suggests the need for further optimization. The findings of this research open new avenues for enhancing predictive maintenance systems.

Keywords: AdaBoost, predictive maintenance, artificial intelligence, failure prediction, TTF

1. Bevezetés

A mesterséges intelligencia (MI) számos iparágat átalakít, és a karbantartás területe sem kivétel. Az érzékelők adatainak folyamatos gyűjtésével, például a hőmérséklet, a nyomás és más működési mutatók alapján az MI jelentősen javíthatja a rendszerek és eszközök felügyeletét és karbantartását. (Syafudin et al.,

2018) A hagyományos, főként ütemezett karbantartásra támaszkodó módszerekkel szemben az MI-alapú megközelítések képesek felismerni azokat a mintákat és rendellenességeket, amelyek működési problémákra vagy meghibásodásra utalhatnak. Ez a képesség nemcsak a hibák előrejelzésének pontosságát javítja, hanem lehetővé teszi a karbantartási ütemtervek optimalizálását is, csökkentve a szükségtelen állásidőt és a kapcsolódó költségeket. Az MI alkalmazása ebben az összefüggésben hozzájárul a környezeti fenntarthatósághoz is. Az optimalizált működés jobb energiahatékonyabb károsanyag-kibocsátást eredményez, ami összhangban van a globális fenntarthatósági célokkal. Ahogy az iparágak tovább fejlődnek, az MI szerepe egyre jelentősebbé válik.

Az MI-alapú karbantartási megközelítések egyre szélesebb körben elterjedtek számos iparágban, különösen azokban, ahol a berendezések megbízható működése kritikus jelentőségű. Az ipari gyártásban, például, az MI-t gyakran alkalmazzák prediktív karbantartásra, amely képes előre jelezni a gépek meghibásodását, így minimalizálva a leállásokat és növelve a termelés hatékonyságát. Egy tanulmány szerint a prediktív karbantartás lehetővé teszi az akár 30%-os költségcsökkentést a hagyományos karbantartási modellekhez képest (Lee et al., 2018). Hasonlóképpen, az energetikai szektorban az MI-t különböző infrastruktúrák, például a szélturbinák vagy elektromos hálózatok állapotának felügyeletére használják, lehetővé téve az eszközök hosszabb élettartamát és a rendszerek hatékonyabb működését (Pinciroli et al., 2022).

Az egészségügyi iparágban az MI alkalmazása szintén jelentős előrelépést hozott a berendezések karbantartásában. A képalkotó diagnosztikai eszközök, például MRI- vagy CT-szkennerek megbízható működése elengedhetetlen a pontos diagnózisok érdekében. Az MI-alapú algoritmusok képesek előre jelezni a kulcsfontosságú alkatrészek elhasználódását, és ezzel csökkenteni a nem tervezett leállások kockázatát (Ben-Bouazza et al., 2023). Az infrastruktúrában, például az építőiparban és a közlekedésben, az MI-t hidak, vasutak és alagutak állapotának felügyeletére használják. Ez nemcsak a biztonságot növeli, hanem hosszú távon jelentős költségmegtakarítást is eredményezhet (McMillan, Varga, 2022). Ezek az iparági példák jól mutatják, hogy az MI-alapú karbantartási megoldások széles körben alkalmazhatók, és komoly előnyöket kínálnak a költséghatékonyság, a megbízhatóság és a fenntarthatóság szempontjából.

2. Kutatási célkitűzés

Jelen cikk az *OmegaSys – Élettartam tervező és meghibásodás előrejelző komplex döntéstámogató rendszer, facility management szolgáltatás kialakításához* című kutatás-fejlesztési projekt megvalósításához kapcsolódik. A kutatási célkitűzés a mesterséges intelligencia- (MI-) alapú előrejelző modellek kidolgozása, implementálása, validálása és tesztelése a létesítménygazdálkodás tervezési, előrejelzési, valamint működtetési és fenntartási folyamataiban. A projekt célja, hogy egy komplex döntéstámogató rendszert hozzon létre, amely lehetővé teszi a létesítmények gépészeti elemei és épületgépészeti rendszerei élettartamának pontos meghatározását már a tervezési szakaszban, valamint a meglévő berendezések optimális, adaptív karbantartási tervének elkészítését.

A kutatás egyik ága a sztochasztikus Markov-modellek (Chen et. al, 2019) és a meghibásodások valószínűségi eloszlásfüggvényei (pl.: Weibull-, exponenciális- és normális eloszlások) paramétereinek meghatározását célozza meg mesterséges intelligencia algoritmusokkal. Az adaptív MI-alapú algoritmusok a létesítményekből gyűjtött adatokat, valamint karbantartási naplókat felhasználva javítják az előrejelzések pontosságát, csökkentve a meghibásodások kockázatát és az üzemeltetési költségeket. A projekt eredményeként létrejövő rendszer a létesítménygazdálkodási szolgáltatók számára új lehetőségeket kínál az üzemeltetési kockázatok csökkentésére, a karbantartási folyamatok optimalizálására és az energiahatékonyság növelésére.

A kutatás másik célja a berendezések élettartamának minél pontosabb előrejelzése mesterségesintelligencia-alapú modellek segítségével, amelyek a létesítmények gépészeti és épületgépészeti elemeinek teljesítményadatait, karbantartási előzményeit és környezeti tényezőit elemzik (Hornýák, 2023). Az MI-alapú megközelítés lehetővé teszi a rendszerek várható élettartamának RUL- (Remaining Useful Life) becslését, ezáltal támogatva a megelőző karbantartási folyamatokat, minimalizálva a meghibásodásokból adódó üzemkimaradásokat és csökkentve a karbantartási költségeket.

E cél érdekében olyan algoritmusok fejlesztése történt, amelyek valós idejű adatgyűjtésen és -elemzésen alapulnak, figyelembe véve a leggyakoribb meghibásodási mintákat, valamint a gépészeti rendszerek működési körülményeit. Az élettartam előrejelzésének pontossága javítja a létesítménygazdálkodás hatékonyságát, miközben hozzájárul a fenntarthatóbb és költséghatékonyabb üzemeltetési gyakorlatok kialakításához.

A kutatás harmadik célja az optimális karbantartási ütemtervek kidolgozása, amelyek az MI-alapú algoritmusok segítségével figyelembe veszik a meghibásodások valószínűségét, a hibaelhárítás és üzemkimaradás költségeit, valamint az erőforrások rendelkezésre állását. Az adaptív ütemezési modellek lehetőséget nyújtanak arra, hogy a karbantartási feladatok időzítése a valós idejű adatok és előrejelzések alapján dinamikusan változzon. Az így kialakított ütemtervek hozzájárulnak a leállások minimalizálásához és a berendezések rendelkezésre állásának maximalizálásához, miközben csökkentik az üzemeltetés teljes költségét. Ez a megközelítés javítja a létesítménygazdálkodás hatékonyságát, növeli az üzleti folyamatok megbízhatóságát, és fenntarthatóbb működést tesz lehetővé.

3. A vizsgált adatok

Ez a tanulmány az **Engine Failures Dataset** (Mologtaliev, 2022) adathalmaz elemzését végzi. A strukturált megközelítés célja, hogy értelmes összefüggéseket tárjon fel, és mesterséges intelligencián alapuló prediktív modellt fejlesszen. A kutatás lépései a következők voltak:

1. **Adatgyűjtés és szemantikai elemzés:**
 - Az adatok beszerzése és feltárása.
 - A dokumentációk áttekintése.
 - Az adatok kezdeti vizsgálata az adattípusok, hiányzó értékek és alapvető statisztikák azonosítására.
2. **Adattisztítás és előfeldolgozás:**
 - Hiányzó értékek vizsgálata és kezelése imputációval.
 - Szélsőértékek kezelése, például levágással, normalizálással vagy standardizálással.
3. **Feltáró adatelemzés:**
 - Az adatokon belüli mintázatok, trendek és kapcsolatok feltárása.
 - Vizualizációs eszközök, hisztogramok, boxplotok és szórási diagramok használata.
4. **Modellfejlesztés:**
 - AdaBoost algoritmuson alapuló prediktív modell létrehozása.
5. **Modellértékelés:**
 - A modell teljesítményének értékelése, például pontosság (accuracy), precizitás (precision), visszahívási arány (recall) és F1-pontszám segítségével.
6. **Eredmények bemutatása:**
 - Az eredmények összefoglalása egy átfogó jelentésben vagy prezentációban.
 - A módszertan, az eredmények és az ajánlások részletezése.

A tanulmány 100 különböző motor folyamatos monitorozásáról szól, amelyeket a használat megkezdésétől egészen a meghibásodásukig követtek nyomon. Minden motor meghatározott körülmények között üzemelt – amelyek eltérhettek például a működési intenzitás, a környezeti feltételek vagy a karbantartási ütemtervek függvényében –, egészen addig, amíg meghibásodtak. Az adatokban szereplő motor típusa nincs meghatározva, és az adatoszlopok mértékegységei sincsenek megadva. Az adathalmaz a következő oszlopokat tartalmazza:

- ID: Ez az oszlop minden monitorozott motor vagy egység egyedi azonosítóját jelöli.
- TTF: Time to Failure, meghibásodásig eltelt idő, amely az elsődleges vizsgált változó. Azt jelzi, hogy mennyi idő van hátra a motor meghibásodásáig. Jelen adathalmaz esetén a működési ciklusok száma.
- s12, s14, s17: Ezek az oszlopok különböző jeleket vagy érzékelői adatokat tartalmaznak, amelyek a motor állapotával kapcsolatosak. Minden jel specifikus adatpontokat nyújt, amelyeket a motor működése során gyűjtöttek.

Az 1. táblázat bemutatja egy véletlenszerűen kiválasztott motorról gyűjtött adatokat.

1. táblázat
Az adathalmaz

ID	TTF	s12	s14	s17
100	4	519,49	8137,6	397
100	3	519,68	8136,5	395
100	2	520,01	8141,05	398

Az adathalmaz 20 631 bejegyzést tartalmaz. Az ID-oszlop értékei 1-től 100-ig terjednek, ami a monitorozott motorok számát jelzi. A TTF, amely ciklusokban van megadva, 0 és 361 között változik, az átlagértéke körülbelül 108. Az adathalmaz nem tartalmaz leírást a ciklus definíciójáról. Motorok tekintetében az alábbi ciklusok szokásosak:

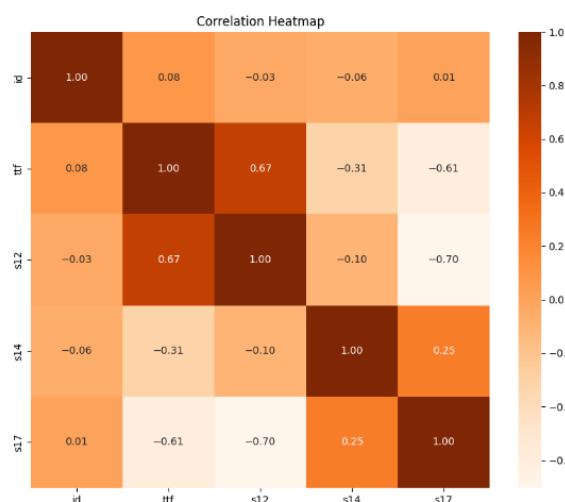
- Üzemeltetési ciklus: Egy motor működésének egy teljes fordulata vagy adott idejű működési szakasza, amely után újraindul vagy folytatódik. Például egy turbina vagy szivattyú meghatározott időközönként végzett működési egysége.
- Karakterisztikus ciklus: A motor által végzett ismétlődő működési fázis, amely egy adott terhelési mintázatot vagy működési körülményt takarhat (például felmelegedés, csúcsterhelés, lehűlés).
- Adatrögzítési ciklus: Az adatgyűjtési folyamat egy periódusa, amely alatt a rendszer rögzíti az állapotváltozásokat, de a pontos időtartama nincs meghatározva az adathalmazban.
- Karbantartási ciklus: Olyan időközönként ismétlődő esemény, amely a motor állapotának ellenőrzését, beavatkozását vagy javítását jelenti.

Mivel az adathalmaz nem tartalmazza a ciklusok időtartamát, nem tudjuk pontosan, hogy egy ciklus mennyi időt ölel fel, de az biztos, hogy a motor működésének és élettartamának vizsgálatára egyfajta ismétlődő egységként van értelmezve. Az adathalmaz rögzítésekor egy motor már meghibásodott (TTF = 0).

A korrelációvizsgálat eredményei alapján a TTF (Time-To-Failure, meghibásodásig hátralévő idő) erős pozitív korrelációt mutat az s12 értékkel (0,67), ami azt sugallja, hogy az s12 növekedésével a rendszer várható élettartama is növekszik. Ezzel szemben az s17-tel erős negatív korreláció (–0,61) figyelhető meg, vagyis ennek az értéknek az emelkedése a meghibásodás gyorsabb bekövetkezésével jár

együtt. A harmadik jel, s14, mérsékelt negatív korrelációt mutat a TTF-fel ($-0,31$), ami arra utal, hogy bár van kapcsolat a két változó között, az nem olyan jelentős, mint az s17 esetében. Ezzel párhuzamosan az s14 és az s17 között mérsékelt pozitív korreláció van ($0,25$), ami arra enged következtetni, hogy az s17 növekedésével az s14 is hajlamos növekedni.

Érdemes kiemelni, hogy az s12 és s17 között erős negatív korreláció tapasztalható ($-0,70$), ami azt jelenti, hogy amikor az s12 értéke növekszik, az s17 hajlamos csökkenni, és fordítva. Ez arra utalhat, hogy ezek a változók ellentétes módon befolyásolják a rendszer állapotát, és esetleg egymással versengő hatásokra utalnak a motor működésében vagy állapotában. Az s12 változó növekedését olyan tényezőként lehet figyelembe venni, amely meghosszabbíthatja a berendezés élettartamát, míg az s17 emelkedése valószínűsíthetően a gyorsabb meghibásodást jelzi előre. Az s14 szerepe összetettebb, és további vizsgálatot igényel annak pontosabb megértéséhez, hogy miként befolyásolja az üzemi ciklusokat. A korrelációs mátrix hőképe az 1. ábrán látható.



1. ábra. Korrelációs analízis

Az egyes jelek statisztikai jellemzőit a 2. táblázat tartalmazza, az adathalmaz oszlopainak leíró statisztikai szintén a 2. táblázatban található.

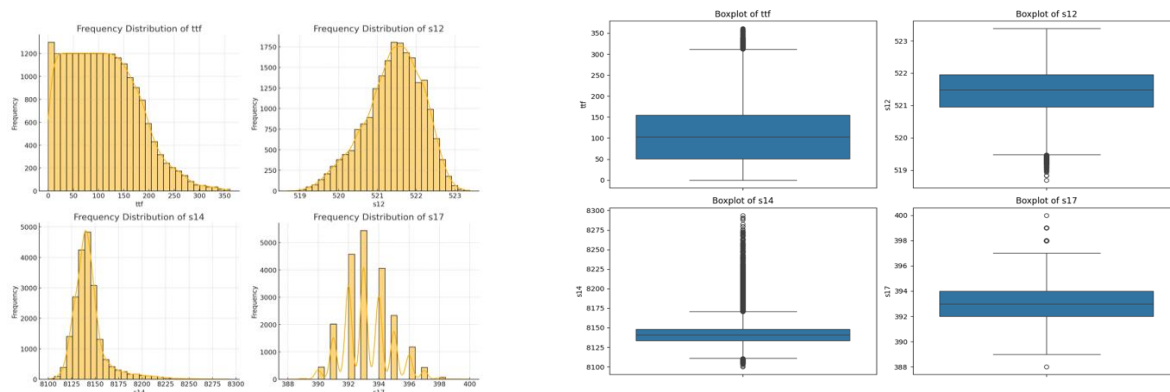
2. táblázat

Az adatokat leíró statisztika

	TTF	s12	s14	s17
Darabszám	20 631	20 631	20 631	20 631
Átlag	107,81	521,41	8143,75	393,21
Szórás	68,88	0,74	19,08	1,55
Min.	0	618,69	8099,94	388,00
Max.	361,00	523,38	8293,72	400,00

A következő megállapítások tehetők: Az adathalmaz 20 631 mezőt tartalmaz az összes oszlopra vonatkozóan, ami egy teljes, hiányzó értékektől mentes adathalmazt jelez. Az s12, s14 és s17 jelek átlag- és mediánértékei közel azonosak, ami ezeknél a méréseknél szimmetrikus eloszlást jelez az értékek közép-pontja körül. Az egyes érzékelő jelek szórása viszonylag kicsi, ami arra utal, hogy az értékek többsége az átlaghoz közel helyezkedik el, következetes érzékelői méréseket sugallva. Az érzékelők viszonylag szűk interkvartilis tartománya (Interquartile Range, rövidítve IQR) azt mutatja, hogy az adatok többsége a mediánhoz közel helyezkedik el, ami megfelelő alapot nyújt a prediktív modellezéshez és a megbízhatóság elemzéséhez.

A jelek értékeinek eloszlását a 2. a) ábra mutatja be, a 2. b) ábra a statisztikai mutatókat ábrázolja.



2. ábra. a) A jelek frekvenciájának eloszlása, b) boxplot diagram

Ez a tanulmány az úgynevezett adaptív boosting (AdaBoost) algoritmust mutat be. Az AdaBoost egy olyan tanulási algoritmus, amely gyenge osztályozókat használ és azok teljesítményét javítja, majd azokat egy erőteljes osztályozóvá egyesíti. Az algoritmus azzal kezdődik, hogy minden tanulási mintának egyenlő súlyt ad. Ezt követően az algoritmus iteratív módon gyenge osztályozókat tanít, és a tanulási minták súlyát a besorolási eredmények alapján állítja be. A helytelenül osztályozott minták súlya növekszik, míg a helyesen osztályozott minták súlya csökken. Ez a folyamat biztosítja, hogy a későbbi tanulók nagyobb figyelmet fordítsanak az összetettebb esetekre. Minden gyenge tanuló hozzájárulását a végső modellhez a hibaaránya határozza meg: minél alacsonyabb a hibaarány, annál nagyobb a súly (alfa). A végső modell a gyenge tanulók előrejelzéseinek súlyozott összege.

Az előfeldolgozási lépés során betöltöttem az adathalmazt, és a TTF-értékek mediánja alapján egy osztályozási küszöböt határozok meg. Bináris címkéket hoztam létre: azok a minták, amelyek TTF-értéke kisebb volt, mint a medián, 1-es (magas kockázat), míg amelyek értéke nagyobb vagy egyenlő volt a mediánnal, -1-es (alacsony kockázat) címkét kaptak. Az osztályozáshoz használt jellemzők mindhárom érzékelőből származó adatot tartalmazták (s12, s14, s17).

Adott:

– Egy tanulási mintahalmaz: $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, ahol x_i a jellemzővektor és y_i a címke (-1 vagy +1).

1. Inicializálás:

– Adjunk minden mintának azonos súlyt: $w_i = 1 / n$, ahol n a minták száma.

2. Iterációk:

Ismételjük a következő lépéseket T-iteráción keresztül:

- a. Egy gyenge tanuló tanítása:
– Tanítsunk egy gyenge tanulót a súlyozott tanulási mintákra.
- b. Hibaarány számítása:
– Számítsuk ki a gyenge tanuló relatív hibáját (ϵ):

$$\epsilon = (\text{súlyozott hibák összegének aránya az összes súlyhoz képest}).$$

$$\epsilon = \sum (w_i * I(h(x_i) \neq y_i)) / \sum w_i$$
 ahol $I()$ egy indikátorfüggvény, amely 1-et ad vissza, ha $h(x_i) \neq y_i$
- c. Gyenge tanuló súlyának meghatározása:
– Számítsuk ki a gyenge tanuló megbízhatóságát (α):

$$\alpha = 0,5 * \ln((1 - \epsilon) / \epsilon)$$
- d. Súlyok frissítése:
– Frissítsük a súlyokat a következő módon:
 Ha $h(x_i) = y_i$ (helyes osztályozás):

$$w_i = w_i * \exp(-\alpha)$$
 Ha $h(x_i) \neq y_i$ (helytelen osztályozás):

$$w_i = w_i * \exp(\alpha)$$
 – Normáljuk a súlyokat, hogy összegük 1 legyen:

$$w_i = w_i / \sum w_i$$
3. Végző modell:
– A végző osztályozó:

$$H(x) = \text{sign}(\sum \alpha_t * h_t(x)), \text{ ahol } t = 1, 2, \dots, T.$$
4. Kimenet:
– A végző osztályozó $H(x)$, amely az összes gyenge tanuló súlyozott eredménye alapján dönt.

Az AdaBoost modellt Pythonban implementáltam a scikit-learn könyvtár AdaBoostClassifier osztályának segítségével. A modell 100 becselővel ($n_estimators = 100$), 1.0 tanulási rátával ($learning_rate = 1.0$) és 42 véletlenszám-generátor állapottal ($random_state = 42$) készült a reprodukálhatóság biztosítása érdekében. A kísérletet egy személyi számítógépen végeztem, amely egy Intel i7 processzorral és 16 GB RAM-mal volt felszerelve.

Az adatokat 60%-os tanulási és 40%-os tesztelési részre osztottam. A modell tanítása után a tesztkészleten értékeltem a teljesítményét. Az osztályozási előrejelzések elvégzése után különböző teljesítménymutatókat számítottak ki, beleértve a konfidenciamátrixot, a pontosságot (accuracy), a precizitást (precision), a visszahívási arányt (recall) és az F1 pontszámot.

A confusionmátrix elemei a következők lettek:

Valós \ Előrejelzett	Pozitív (Előrejelzett)	Negatív (Előrejelzett)
Pozitív (Valós)	3530 (True Positive, TP)	1081 (False Negative, FN)
Negatív (Valós)	548 (False Positive, FP)	3094 (True Negative, TN)

Kiszámítottam a teljesítménymutatókat:

1. Pontosság:

A helyes előrejelzések aránya az összes előrejelzéshez képest:

$$\text{Pontosság} = (TP + TN) / (TP + TN + FP + FN) = 80,29\%$$

2. Precizitás:

A helyesen azonosított pozitív esetek aránya az összes pozitív előrejelzéshez képest:

$$\text{Precizitás} = TP / (TP + FP) = 86.56\%$$

3. Visszahívás vagy Érzékenység:

A helyesen azonosított pozitív esetek aránya az összes valós pozitív esethez képest:

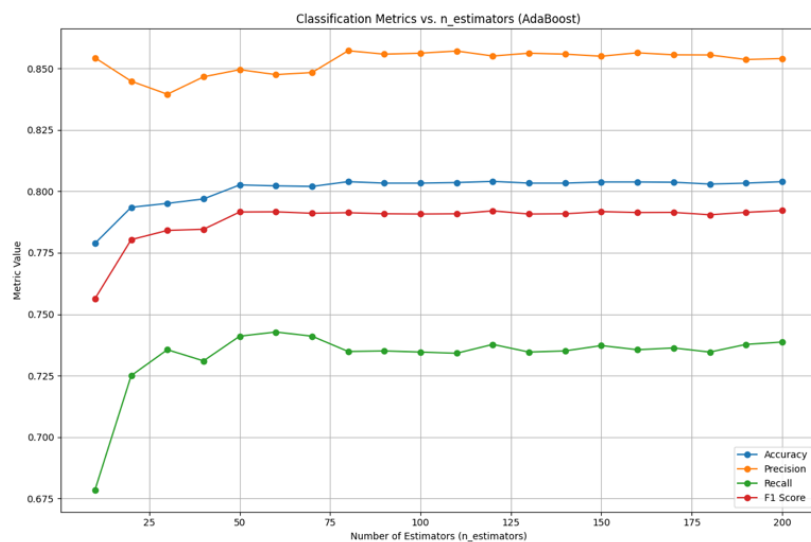
$$\text{Visszahívás} = \text{TP} / (\text{TP} + \text{FN}) = 76,55\%$$

4. F1 pontszám:

A precizitás és visszahívás harmonikus átlaga:

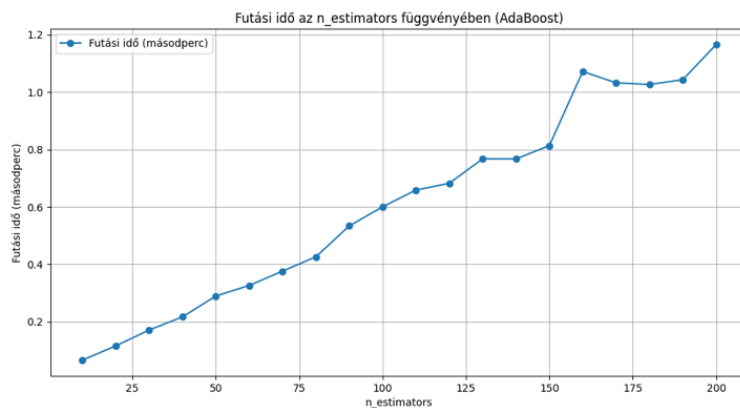
$$\text{F1} = 2 \cdot \text{Precizitás} \cdot \text{Visszahívás} / (\text{Precizitás} + \text{Visszahívás}) = 81,62\%$$

Megvizsgáltam az algoritmus hiperparamétereit. Az osztályozók száma amikor elérte a 80-at, onnan a teljesítménymutatók érdemben nem változtak. Ezt a 3. ábra mutatja.



3. ábra. A hiperparaméterek vizsgálata

A gyenge tanulók számának növelése a futási időt lineárisan növelte. Ehhez egy külön kísérletet végeztem és az eredményeit ábrázoltam.



4. ábra. A gyenge tanulók számának hatása a futási időre

4. Értékelés

A modell helyesen osztályozza az esetek nagy részét. A pozitív előrejelzések közül sok helyes, tehát a téves pozitívok aránya alacsony. Az érzékenység azt mutatja, hogy a modell a valós pozitív esetek körülbelül háromnegyedét felismeri, de a hamis negatívok (elmulasztott pozitív esetek) száma még jelentős.

Az F1 pontszám kiegyensúlyozottsági mutató, amely a precizitás és a visszahívás közötti kompromisszumot értékeli. Ez az érték azt mutatja, hogy a modell jól teljesít az általános osztályozási feladatokban, különösen, ha a precizitás fontosabb, mint az érzékenység.

Javítási javaslatok:

1. Ha az érzékenység növelése fontos (például, ha a pozitív esetek elmulasztása kritikus), a modell további finomhangolására van szükség, például a döntési küszöb csökkentésével.
2. Ha a téves pozitívok minimalizálása lényeges, a precizitás további növelése érdekében a hamis pozitív eseteket lehet célzottan csökkenteni, például pontosabb jellemzők kiválasztásával vagy a modell paramétereinek optimalizálásával.

5. Összefoglalás

Az adatvezérelt megközelítések, mint például a jelen tanulmányban alkalmazott módszer egyik fő előnye, hogy nem igényelnek kiterjedt előzetes ismereteket az autómotorok mechanikájáról. Ez lehetővé teszi, hogy az elemzés kizárólag az adatok alapján történjen, függetlenül attól, hogy az egyes adatok pontosan mit jelentenek vagy milyen részleteket tartalmaznak. Azonban ezeknek a megközelítéseknek a sikere a modell képzéséhez és teszteléséhez használt szenzoradatok megbízhatóságán és pontosságán múlik.

Ez a tanulmány elméleti megközelítést mutatott be. Egy AdaBoost osztályozási modellt alkalmaztak az autómotorok meghibásodási idejének előrejelzésére a szenzorok által mért adatok alapján. A modell hozzávetőlegesen 80%-os pontosságot ért el, ami az AdaBoost számára viszonylag magas teljesítményt jelent. A 86%-os precizitás azt mutatja, hogy a modell megbízhatónak tekinthető. A 76,5%-os visszahívási arány (recall) azt jelzi, hogy a modell képes az „alacsony kockázatú” esetek többségét azonosítani. Bár ez egy jó teljesítmény, arra is rámutat, hogy van még lehetőség a valódi pozitív esetek pontosabb azonosításában.

Fontos megjegyezni, hogy ez a tanulmány hipotetikus. Ahhoz, hogy az itt bemutatott megközelítés mérnöki megoldásként validálható legyen, megbízható adatokra van szükség – ideértve a pontos szenzoradatokat, a megfelelő mintavételi rátákat, valamint részletes információkat a mért dimenziók és a motor üzemi körülményeinek interakciójáról. A valós üzemi körülményekkel és a szenzoradatok validációja nélkül az eredmények elméletinek tekinthetők.

Amennyiben a modell előrejelzéseit megbízható adatokkal validálják, lehetőség nyílik a megelőző karbantartási ütemtervek javítására, a váratlan motorhibák csökkentésére és az üzemeltetési hatékonyság növelésére. Azáltal, hogy pontosan előre jelezzük a meghibásodásokat, a karbantartási erőforrásokat hatékonyabban lehetne allokálni a magasabb kockázatú motorokra, optimalizálva ezzel a karbantartási erőfeszítéseket és költségeket.

6. Köszönetnyilvánítás

A cikkben ismertetett kutatómunka a Nemzeti Kutatási Fejlesztési és Innovációs Hivatal 2020-1.1.2-PIACI-KFI-2020-00147 számú pályázati támogatásával valósult meg. A szerző köszöni a támogatást.

Irodalom

- [1] Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 18, 20–23. <https://doi.org/10.1016/j.mfglet.2018.09.002>
- [2] Pinciroli, L., Baraldi, P., Ballabio, G., Compare, M., & Zio, E. (2022). Optimization of the Operation and Maintenance of renewable energy systems by Deep Reinforcement Learning. *Renewable Energy*, 183, 752–763. <https://doi.org/10.1016/j.renene.2021.11.052>
- [3] Ben-Bouazza, F. E., Manchadi, O., Dehbi, Z. E. O., Rhalem, W., & Ghazal, H. (2022). Machine learning based predictive maintenance of pharmaceutical industry equipment. *International Conference on Advanced Intelligent Systems for Sustainable Development*, pp. 497–514. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35245-4_46
- [4] McMillan, L., & Varga, L. (2022). A review of the use of artificial intelligence methods in infrastructure systems. *Engineering Applications of Artificial Intelligence*, 116, 105472. <https://doi.org/10.1016/j.engappai.2022.105472>
- [5] Hornyák, O. (2024). From data to decisions: AdaBoost-based approach for remaining useful life estimation. *25th International Carpathian Control Conference (ICCC)*, pp. 1–7. IEEE., <https://doi.org/10.1109/ICCC62069.2024.10569647>
- [6] Syafrudin, M., Alfian, G., Fitriyani, N. L., & Rhee, J. (2018). Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. *Sensors*, 18 (9), 2946. <https://doi.org/10.3390/s18092946>
- [7] Mologtaliev, A. (2022). *Engine Failure Dataset*. Available online: <https://www.kaggle.com/datasets/m0ntecarl0/engine-time-to-failure> (accessed on 28 July 2024).
- [8] Chen, Z., Li, Y., Xia, T., & Pan, E. (2019). Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy. *Reliability Engineering & System Safety*, 184, 123–136. <https://doi.org/10.1016/j.ress.2017.09.002>
- [9] Hornyák, O. (2023). LSTM modellek felhasználása a meghibásodások előrejelzésében: Leveraging LSTM Models for Failure Prediction. *2023 Energetika-Elektrotechnika–Számítástechnika és Oktatás Multi-konferencia*, pp. 128–133.