

Predikcia hodnoty rozsevu aglomerátu pomocou LSTM modelu

Dávid Firda, Jonáš Sirotnák
Technická Univerzita v Košiciach
Fakulta elektrotechniky a informatiky

david.firda@student.tuke.sk, jonas.sirotnak@student.tuke.sk

Abstract—Tento článok skúma využitie neurónových sietí s dlhodobou krátkodobou pamäťou (LSTM) na predikciu veľkosti častíc aglomerátu. Tradičné metódy predikcie a kontroly veľkosti častíc čelia výzvam, ako je variabilita surovín a dynamika výrobných podmienok. Využitím schopnosti LSTM sietí modelovať časové závislosti a zachytiť dynamiku sekvenciálnych dát, táto štúdia demonštruje vhodnosť LSTM sietí v kontexte časovo meniacich sa výrobných procesov. Integrácia pokročilých analytických techník do priemyselných prostredí je ukázaná ako prostriedok na zlepšenie rozhodovania v procesoch a celkovej efektivity výroby. Hodnotenie LSTM modelov pri predikcii veľkosti častíc v rámci procesu aglomerácie poskytuje cenné poznatky pre optimalizáciu výrobného výkonu a kontroly kvality.

Index Terms—Prediction, LSTM neural network, recurrent neural networks, data processing

I. ÚVOD

V súčasnej priemyselnej praxi je optimalizácia výrobných procesov, redukcia nákladov a zlepšenie kvality produktov kľúčovou prioritou. Jedným z významných procesov v oceliarskom priemysle je výroba aglomerátu, ktorá priamo ovplyvňuje efektívnosť a ekonomickú výkonnosť celkového výrobného procesu. Kritickým faktorom pre zabezpečenie kvality aglomerátu je jeho granulometrické zloženie, známe ako rozsev. Tradičné metódy predikcie a kontroly rozsevu čelia výzvam, ako sú variabilita surovín a dynamika výrobných podmienok.

Tento report sa zaoberá využitím moderných metód strojového učenia, konkrétne sietí LSTM [1], [2] (Long Short-Term Memory), na predikciu hodnoty rozsevu aglomerátu. Vďaka schopnosti modelovať časové závislosti a zachytiť dynamiku sekvenciálnych dát, sú LSTM siete ideálne pre aplikáciu v kontexte časovo variabilných výrobných procesov.

Cieľom tohto reportu je hodnotiť efektívnosť LSTM modelov pri predikcii granulometrického rozloženia v rámci výrobného procesu aglomerátu. Taktiež poskytuje prehľad o tom, ako integrácia pokročilých analytických techník do priemyselného prostredia môže prispieť k lepším procesným rozhodnutiam a zvýšiť celkovú efektivitu výroby.

II. DÁTA

Dáta z výroby aglomerátu v spoločnosti US Steel v Košiciach predstavujú cenný zdroj informácií o procese

výroby aglomerátu. Tieto dáta zahŕňajú rôzne merania zo senzorov umiestnených na špekacom páse a taktiež zaznamenávajú hodnoty rozsevu 0-5 milimetrov. Dataset pozostáva z meraní uskutočnených v 30-minútových intervaloch počas štyroch rokov z jedného špekacieho pásu.

Dôležité je tiež poznamenať, že v kontexte výroby aglomerátu je hodnota rozsevu dôležitým ukazovateľom kvality výrobku. Ak je hodnota rozsevu menšia ako 12, aglomerát sa považuje za výrobok dobrej kvality. Avšak ak hodnota rozsevu prekročí tento limit, kvalita aglomerátu sa znižuje, čo môže mať negatívny vplyv na výsledný produkt a spôsobiť ďalšie problémy v procese výroby.

Tieto dáta majú zásadný význam pre monitorovanie a riadenie výrobného procesu, umožňujú analyzovať a predvídať vývoj výroby a identifikovať prípadné anomálie alebo nedostatky v procese výroby. Ich správne využitie môže viesť k zlepšeniu výrobných výkonov, optimalizácii procesov a znižovaniu nákladov na údržbu a opravy.

A. Predspracovanie dát

Naše predspracovanie údajov zahŕňa kroky aplikované na dva súbory dát. Prvý súbor (DT1) obsahuje merania aglomerátu v 30-minútových intervaloch, zatiaľ čo druhý súbor (DT2) obsahuje manuálne zaznamenané merania o 11:00 a 23:00 hodine. Účelom je pripraviť údaje na analýzu a modelovanie. Začíname úpravou časových údajov, konvertujeme ich na dátumové objekty a zarovnávame ich podľa zadanych časových intervalov.

V ďalšej časti predspracovania sme upravili časové intervaly v súbore DT2, nastavili sme ich na konkrétne hodiny a v prípade potreby posunuli na predchádzajúci deň. Potom sme spracovali súbor DT1, odstránili sme neplatné hodnoty a irelevantné údaje.

Nakoniec sme zlúčili oba súbory podľa zhodných časových intervalov a vyčistili výsledný dataset od nepotrebných stĺpcov. Tieto kroky zabezpečujú pripravenie dát pre ďalšiu analýzu a spracovanie neurónovou sieťou.

B. Analýza dát

Pri analýze distribúcie hodnôt "Rozsev 0-5" je zrejmé, že väčšina hodnôt sa zoskupuje okolo centrálnej hodnoty

a sleduje normálne rozdelenie s vrcholom okolo 10. Avšak, pozorujeme aj výskyt extrémnych hodnôt, ktoré dosahujú až k 40. Tieto vysoké hodnoty sú výsledkom špecifických podmienok, čo naznačuje, že nie sú náhodné, ale sú ovplyvnené konkrétnymi faktormi. Identifikácia a analýza týchto špecifických podmienok sú kľúčové pre porozumenie dôvodov výskytu týchto vysokých hodnôt.

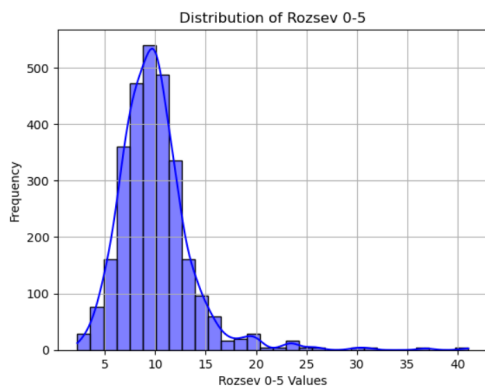


Fig. 1. Distribúcia spracovaných dát.

Tieto informácie sú kľúčové pre presnú charakterizáciu štatistických vlastností dátového súboru. Tieto informácie sú podstatné pre zistenie či získané dáta sú relevantné a presné, čo je zásadné pre ich efektívne využitie v ďalších krokoch.

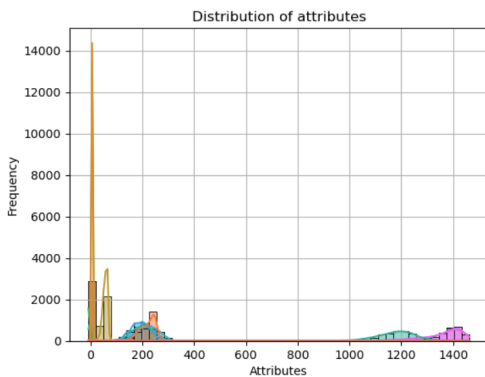


Fig. 2. Distribúcia spracovaných dát.

Ďalší graf ukazuje výkyvy, ktoré môžu zodpovedať určitým vonkajším vplyvom alebo vnútorným procesom, ktoré ovplyvňujú cieľovú premennú. Môže ísť o dáta zaznamenávané v rôznych časových obdobiach, kde vysoké špičky môžu indikovať mimoriadne udalosti alebo zmeny v podmienkach.

Dôležité je všimnúť si, že dáta sú rozložené v pomerne rovnomerných intervaloch, čo naznačuje, že merania boli vykonávané systematicky a konzistentne. Tento graf môže byť užitočný na identifikáciu trendov, cyklov alebo na ďalšie štatistické analýzy zamerané na porozumenie dynamiky cieľovej premennej v zadanom časovom rozmedzí.

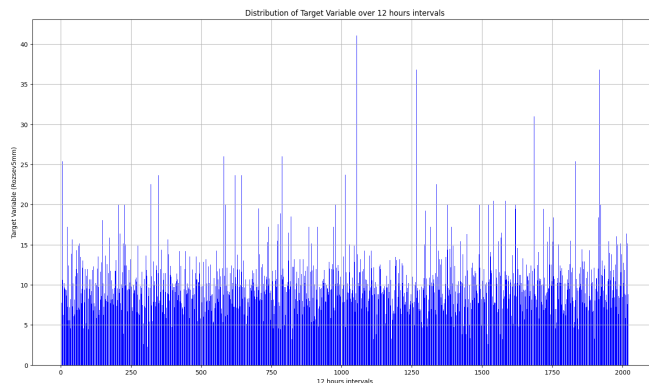


Fig. 3. Distribúcia cieľovej hodnoty v 12 hodinových intervaloch.

C. Spracovaný dataset

Výsledný spracovaný dataset obsahuje už len samotné relevantné dáta z meraní o 11:00 a 23:00 hodine, bez žiadnych dátumových objektov. Teda samotný dataset obsahuje hodnoty všetkých požadovaných atribútov a presnej hodnoty rozsevu. Takto upravený je pripravený na posledné predspracovanie, ktoré je potrebné pred vstupom do neronovej siete.

D. Rozdelenie dát pre vstup do neuronovej siete

V prvom rade sa vstupný dataset (IDT) prevedie na typ float, aby sa zabezpečila číselná konzistencia údajov. Následne pre zbavenie sa prípadných chýbajúcich hodnôt sa zbavujeme hodnôt typu NaN, čím sa zabezpečí integrita a úplnosť údajov na ďalšie spracovanie. Potom vstupné dáta rozdeľujeme na polia vstupných prvkov a cieľových hodnôt. Proces rozdelenia údajov zahŕňa vytvorenie tréningovej (ktorá obsahuje náhodne vybraných 80% dát) a dočasnej množiny (ktorá obsahuje zvyšných 20% dát). Následne sa dočasná množina ďalej rozdelí na validačnú a testovaciu množinu s pomerom 1:1.

Všetky množiny sa potom transformujú pomocou Min-MaxScaler, aby sa zabezpečilo, že každá funkcia má hodnoty v rozsahu medzi -1 a 1, čo je dôležité pre mnohé algoritmy strojového učenia vrátane neurónových sietí.

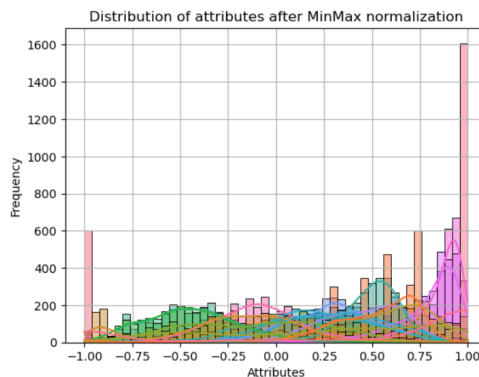


Fig. 4. Distribúcia dát po normalizácii.

Tieto množiny sú následne konvertované na tenzory kvôli kompatibiliti pre vstup do neurónových sietí. Táto konverzia zahŕňa špecifikáciu dátového typu ako np.float32 a zmenu tvaru cieľových tenzorov tak, aby mali tvar (batch_size, 1), ktorý zodpovedá očakávanému formátu na tréning modelu. Trenovacia a validačná množina je pomocou triedy DataLoader rozdelená na menšie dávky pre uľahčenie spracovania dát počas fáz tréningu a validácie, čím sa zvyšuje výpočtová efektívnosť a stabilita tréningu modelu.

III. LONG SHORT-TERM MEMORY SIET'

LSTM, alebo Dlhodobá krátkodobá pamäť, predstavuje pokročilú formu rekurentnej neurónovej siete, ktorú vyvinuli Hochreiter a Schmidhuber [1]. Tento typ siete je ideálny na predikcie sekvencií a exceluje v identifikácii dlhodobých vzťahov v dátach. Používa sa v aplikáciách zahrňujúcich časové rady a rôzne sekvencie. Významnou prednosťou LSTM je jej schopnosť rozumieť vzájomným závislostiam v dátach, čo je prínosné aj pri riešení komplexných úloh, ako sú preklad textu, rozpoznávanie hlasu a podobné zložité úlohy.

A. Čo je to LSTM

Tradičné RNN majú problém s dlhodobými závislosťami kvôli jedinému skrytému stavu, ktorý prechádza v čase a sťažuje sieťam učenie. Tento problém adresuje LSTM prostredníctvom pamäťovej bunky, ktorá dokáže uchovávať informácie na dlhšie obdobie. Vďaka tomu sa siete LSTM dokážu naučiť dlhodobé vzťahy v sekvenciách dát. LSTM môžu byť tiež efektívne kombinované s inými architektúrami neurónových sietí.

Pamäťová bunka LSTM je riadená tromi bránami: vstupnou (input gate), zabudlivou (forget gate) a výstupnou (output gate). Tieto brány umožňujú selektívne rozhodovať, ktoré informácie sa majú pridať do pamäťovej bunky, ktoré sa z nej majú odstrániť a ktoré sa majú z nej vypustiť. Tento mechanizmus umožňuje LSTM sieťam efektívne spracovávať informácie v sekvenciách dát a selektívne udržiavať alebo eliminovať informácie počas ich prechodu sieťou. Veľkosť okna (window) v LSTM modeloch definuje, ako veľa predchádzajúcich krokov (časových intervalov) model berie do úvahy pri učení závislostí v dátach.

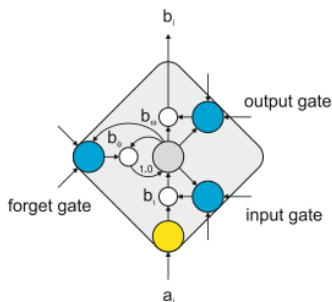


Fig. 5. LSTM pamäťová bunka s bránami [2].

Výsledná pamäťová bunka je znázornená na obr.3, kde sme zahrnuli dve modifikácie pôvodnej jednotky LSTM navrhovanej v [3] a [4].

B. Prečo práve LSTM sieť

Pri výbere modelu na predikciu hodnoty Rozsevu je nevyhnutné zohľadniť niekoľko faktorov. Medzi ne patrí štruktúra dostupných dát, charakteristika samotného časového radu a požiadavky na presnosť predikcií. Modely LSTM (Long Short-Term Memory) sa často javia ako vhodná voľba pre takéto úlohy z viacerých dôvodov:

- **Schopnosť zachytiť dlhodobé závislosti:** LSTM siete excelujú v udržiavaní a aktualizovaní informácií z minulosti v dlhodobej pamäti. Táto vlastnosť je dôležitá pri predikovaní hodnôt v časových radoch, kde sú dlhodobé vzory a závislosti kľúčové.
- **Spracovanie sekvenčných dát:** LSTM modely sú efektívne vo vyhodnocovaní sekvenčných dát, ako sú časové rady. Sú navrhnuté tak, aby sa dokázali učiť z postupnosti dát a identifikovať v nich vzory.
- **Flexibilita:** Parametre LSTM modelov sú nastaviteľné a optimalizovateľné v závislosti od konkrétnych podmienok. To umožňuje prispôbenie modelu špecifickým požiadavkám a charakteristikám dát.
- **Spracovanie zložitých dátových vzorov:** V prípade, že dáta obsahujú viacero vzorov s rôznymi časovými štruktúrami, LSTM modely majú tendenciu byť účinnéjšie pri zachytávaní tejto zložitosti v porovnaní s tradičnými modelmi, ako je napríklad ARIMA.

Tieto vlastnosti robia z LSTM modelov silných kandidátov na predikciu hodnôt Rozsevu, avšak je dôležité vykonať aj ďalší dôkladný prieskum a experimentáciu s modelmi, aby sa dosiahla optimálna predikčná presnosť.

C. Výhody a nevýhody LSTM

Výhody LSTM [5]:

- Siete LSTM sú schopné zachytiť dlhodobé závislosti v dátach vďaka pamäťovej bunke, ktorá umožňuje uchovávanie informácií na dlhú dobu.
- V tradičných RNN dochádza k problému miznúcich a explodujúcich gradientov pri tréningu modelov na dlhých sekvenciách. Siete LSTM tento problém riešia pomocou bránového mechanizmu, ktorý selektívne zapamätáva alebo zabúda informácie.
- LSTM umožňuje modelu zachytiť a pamätať si dôležitý kontext, aj keď medzi relevantnými udalosťami v sekvencii existuje významná časová medzera.

Nevýhody LSTM [5]:

- V porovnaní s jednoduchšími architektúrami, ako sú dopredné neurónové siete, sú siete LSTM výpočtovo náročnejšie.
- Tréningovanie sietí LSTM je časovo náročnejšie v porovnaní s jednoduchšími modelmi kvôli ich výpočtovej zložitosti. Často si vyžaduje viac dát a dlhší čas tréningu, aby dosiahlo vysoký výkon.

D. Štruktúra modelu

Náš model LSTM je navrhnutý pomocou knižnice PyTorch [6] a pozostáva z niekoľkých komponentov na spracovanie sekvenčných údajov.

Metóda forward implementuje tok dát cez model. Začína sa inicializáciou počiatočných skrytých stavov pre vrstvu LSTM. Táto architektúra modelu sa skladá predovšetkým z vrstvy LSTM a následných plne prepojených vrstiev na generovanie konečných výstupných predikcií. Počnúc vstupnou vrstvou, ktorá prijíma dáta, cez LSTM vrstvu, ktorá spracúva a udržiava informácie v pamäti, až po výstupné vrstvy, ktoré generujú predikcie. Prvá vrstva je LSTM vrstva s parametrami definovanými ako počet vstupných dát, veľkosť pamäte, a hodnotou dropout pre regularizáciu na zvýšenie robustnosti učenia.

Následne nasleduje postupné prepojenie s plne prepojenými (FC) vrstvami, ktoré zodpovedajú za transformáciu a zmenšovanie dimenzií dát. Každá z týchto vrstiev je doplnená o dropout, čo pomáha proti pretrénovaniu. Celková štruktúra zahŕňa spojenie týchto vrstiev pomocou relu aktivačných funkcií a lineárnej aktivačnej funkcie, pre výstupnú vrstvu, ktorá vracia predikcie.

E. Parametre modelu

Parametre nášho modelu LSTM sú kritické pre úspešné tréningovanie a predikciu. Na základe veľkosti vstupných dát sme nastavili dimenziu vstupu na počet stĺpcov v našich dátach. Skrytý rozmer (hidden state), ktorý sme nastavili na hodnotu 150, určuje veľkosť pamäte LSTM vrstvy. Ďalším dôležitým parametrom je počet vrstiev LSTM, ktoré sme nastavili na 3, čo ovplyvňuje hĺbku a zložitosť našej siete.

Model sme inicializovali s dropout hodnotou 0,1, čo pomáha proti pretrénovaniu tým, že dočasne vynuluje váhy niektorých neurónov počas tréningovania. Použili sme Mean Squared Error (MSE) ako našu chybovú funkciu, ktorá meria rozdiel medzi predikciami a skutočnými hodnotami. Pre tréningovanie sme zvolili optimalizačný algoritmus Adam s learning rate nastaveným na 0,005, ktorý pomáha s aktualizáciou váh modelu počas spätného šírenia chyby. Tieto parametre sme vybrali s ohľadom na optimalizáciu tréningovania a dosiahnutie najlepších výsledkov pri predikcii hodnôt našich časových radov.

IV. VYHODNOTENIE

V aktuálnej evaluácii modelu sa prejavujú dobré výsledky, čo potvrdzujú nielen metriky strát, ale aj vizuálne porovnania predikcií so skutočnými hodnotami. Strata pri tréningovaní dosiahla hodnotu 0.814, zatiaľ čo strata na validačnej množine je ešte nižšia, čo je 0.260. Tieto hodnoty naznačujú, že model dosahuje dobrú presnosť a účinnosť, čo sa týka učenia a generalizácie na nové dáta. Nízka validačná strata, v porovnaní so stratou tréningovania,

naznačuje, že model nie je preučení a dobre generalizuje, čo je pozitívnym znakom pre jeho schopnosť správne fungovať na nových dátach.

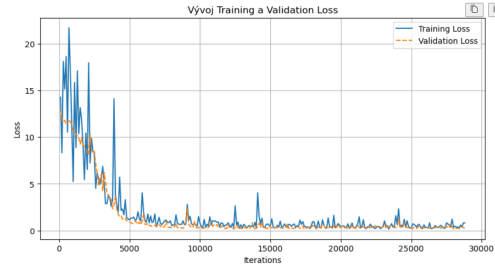


Fig. 6. Vývoj tréningových a validačných chýb.

Z grafu je zjavné, že obidve krivky strát rýchlo klesajú v úvodných fázach tréningovania, čo naznačuje, že model efektívne sa učí z dát. Po tomto rýchlom poklese, krivky strát sa stabilizujú a vykazujú len malé výkyvy.

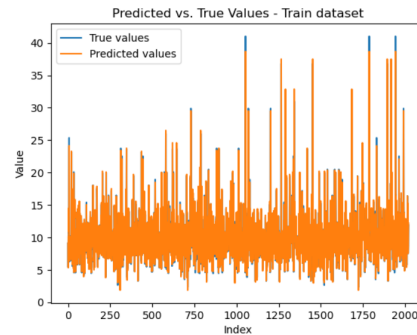


Fig. 7. Predikcia hodnôt na tréningovej množine.

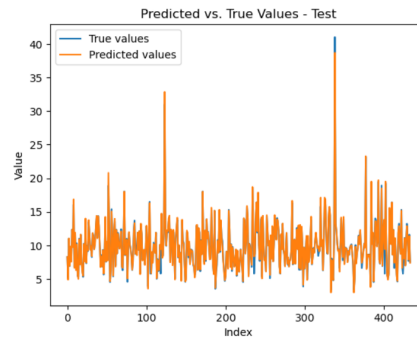


Fig. 8. Predikcia hodnôt na testovacej množine.

Z ďalších grafov je zjavné, že model dokáže zachytiť trend skutočných hodnôt, ale v mnohých prípadoch sú predpovedané hodnoty nižšie a pokrývajú menší rozsah hodnôt v porovnaní s modrou čiarou teda skutočnými hodnotami, ktoré vykazujú výraznejšie vrcholy a hlboké údolia. To naznačuje, že model má problém s presnou predikciou extrémnych hodnôt alebo väčších výkyvov v dátach.

V rámci evaluácie modelu sme použili rôzne metriky na posúdenie výkonnosti nášho modelu. Tieto metriky poskytujú pohľad na rôzne aspekty presnosti a účinnosti modelu pri predpovedaní cieľových hodnôt. Priemerný rozdiel medzi predpovedanými a skutočnými hodnotami a teda MSE nášho modelu bol vypočítaný na 0.2640. Podobne sa zistila aj priemerná absolútna chyba (MAE) 0.2332, ktorá meria priemerný absolútny rozdiel medzi predpovedanými a skutočnými hodnotami, čo poskytuje ďalší pohľad na presnosť modelu.

Okrem toho sme hodnotili kompatibilitu modelu pomocou metriky R-squared (R2), pričom sme získali hodnotu 0.9800. R2 kvantifikuje podiel rozptylu závislej premennej, ktorý sa dá vyjadriť z nezávislých premenných. Vyššia hodnota R2 bližšie k 1 znamená lepšiu zhodu modelu s údajmi.

Taktiež sme použili aj metriku presnosti, čo viedlo k hodnote presnosti 0.9282. Táto metrika vyhodnocuje podiel predpovedí, ktoré sa nachádzajú v určenom rozsahu skutočných hodnôt, a ponúka praktický pohľad na výkonnosť modelu, najmä v scenároch, kde je požadovaná určitá úroveň presnosti.

Nakoniec sme zaviedli našu metriku presnosti, ktorá na rozdiel od tradičných metrick presnosti, ktoré sa zameriavajú výlučne na správne predpovede, naša metrika presnosti zohľadňuje bias predpovedí vzhľadom na skutočné hodnoty.

Táto metrika presnosti počíta presnosť na základe absolútneho rozdielu medzi súčtom skutočných hodnôt a súčtom predikovaných hodnôt vydeleného celkovým počtom vzoriek. Tento prístup poskytuje mieru toho, ako dobre model predpovedá celkový trend údajov, pričom zohľadňuje potenciálne skreslenia v predpovediach.

Zahrnutím týchto metrick môžeme celkovo vyhodnotiť predikčnú výkonnosť modelu a poukázať na jeho silné stránky a oblasti, ktoré je potrebné zlepšiť. Usmerňujú jeho ďalšie zdokonaľovanie alebo použitie v reálnych scenároch.

ZÁVER

Využitie LSTM neurónových sietí v predikcii hodnoty rozsevu aglomerátu v oceliarskom priemysle predstavuje perspektívnu metódu, ktorá môže viesť k významnému zlepšeniu efektivity výrobných procesov a kontrole kvality.

Model LSTM preukázal schopnosť efektívne zachytávať dlhodobé závislosti v časovo meniacich sa výrobných procesoch a presne predikovať hodnoty rozsevu aglomerátu. Vysoká presnosť a generalizačná schopnosť modelu naznačuje ich potenciál v oblasti optimalizácie výrobných procesov a zlepšenia kontroly kvality v

priemyselných aplikáciách.

Tieto pozitívne výsledky podporujú dôležitosť a efektivnosť využitia pokročilých technológií strojového učenia, ako sú aj LSTM siete, v priemyselných odvetviach s cieľom dosiahnuť zvýšenú efektivitu v rámci výrobných procesov.

REFERENCES

- [1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [2] M. Sundermeyer, R. Schlüter, and H. Ney, "Lstm neural networks for language modeling," in *Interspeech*, 2012. [Online]. Available: <https://api.semanticscholar.org/CorpusID:18939716>
- [3] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, Oct 2000.
- [4] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with lstm recurrent networks," *Journal of Machine Learning Research*, vol. 3, pp. 115–143, 2002.
- [5] Aakarshachug, "Deep learning: Introduction to long short-term memory," <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>, 2023, accessed: 2024-05-06.
- [6] PyTorch, "torch.nn.LSTM documentation," <https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html>, 2023, accessed: 2024-05-06.