

# INTEGRATING PCA WITH DEEP LEARNING MODELS FOR STOCK MARKET FORECASTING

Členovia skupiny: Zuzana Orinčáková & Richard Póša

Zdroj: <https://doi.org/10.1016/j.jksuci.2024.102162>

## Abstrakt

- V pôvodnej štúdii bola analýza hlavných komponentov (PCA) použitá na filtrovanie kľúčových ukazovateľov na predpovedanie cien tureckých akcií, pričom sa riešila vysoká volatilita trhu, nelinearita a šum. Vyhodnotili sa štyri modely hlbokého učenia, pričom hybridný model PCA-LSTM-CNN vykazoval vyššiu presnosť vďaka využitiu dlhodobých aj krátkodobých vzorov v akciových údajoch. Pri našej replikácii štúdie sme identifikovali oblasti, ktoré je potrebné zlepšiť, a našli určité nezrovnalosti v pôvodnej metodike, najmä v oblasti predbežného spracovania údajov a výpočtov ukazovateľov. Opravou ukazovateľov a optimalizáciou normalizácie údajov sme sa zamerali na zlepšenie vstupných dát.

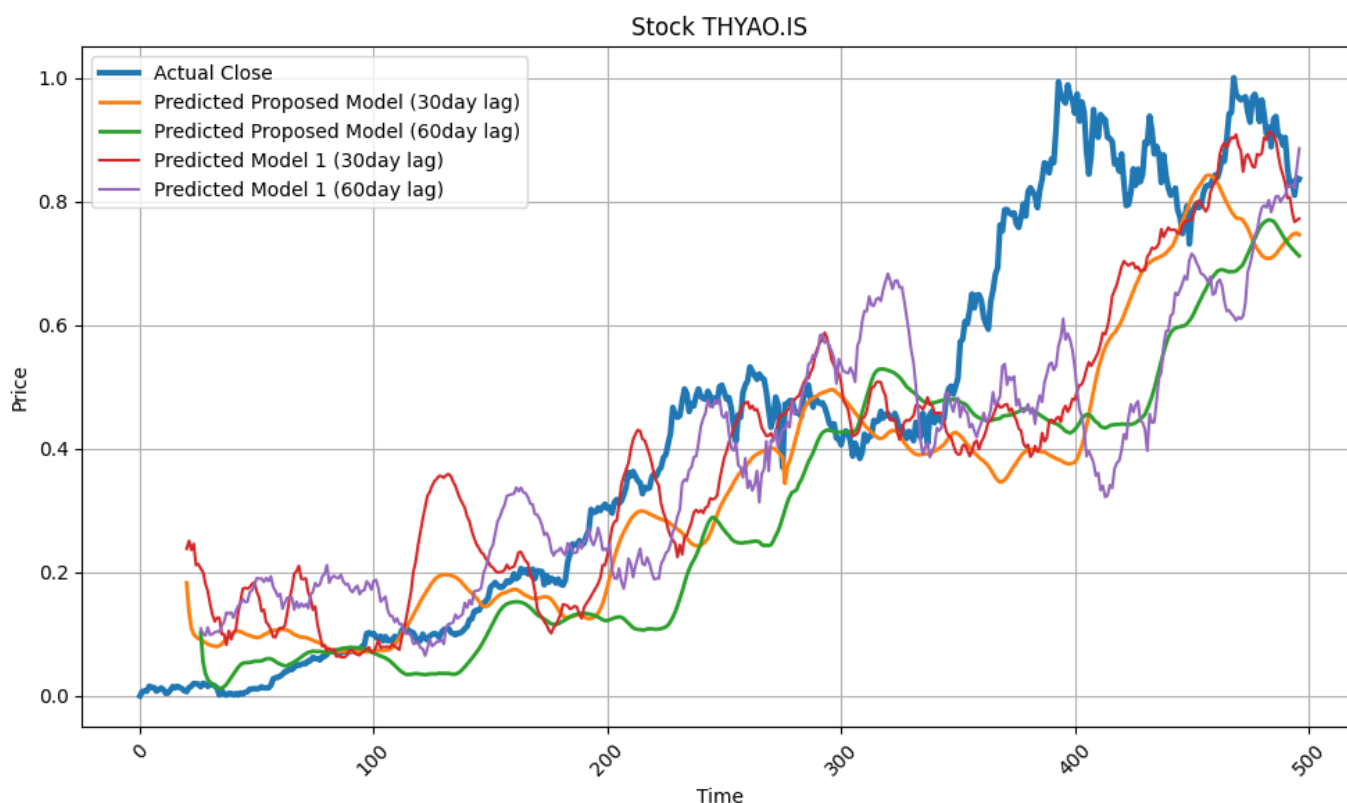
## Naše zmeny oproti pôvodnej metodológii

- Predbežné spracovanie a normalizácia údajov:** V pôvodnej štúdii sme zistili nezrovnalosti pri spracovaní údajov, najmä vo fázach normalizácie a štandardizácie. Tieto procesy sme upravili s cieľom zabezpečiť konzistentné formátovanie vstupných údajov. A to min-max normalizovaním už pripravenej trénovacej, validačnej a testovacej množiny dát.
- Oprava výpočtov technického indikátora:** Pôvodné výpočty niektorých technických ukazovateľov, ako napríklad psychologického línie (PLine), boli nesprávne alebo nemali zadané parametre. Tieto výpočty sme skontrolovali, opravili a nezadané parametre nahradili štandardnými.
- Rozsah použitia a zovšeobecnenie:** Zvážili sme obmedzenia pôvodnej štúdie, ktorá sa zameriavala na jeden trh, navrhujeme rozšíriť model na ďalšie trhy, aby sme posúdili širšiu uplatniteľnosť. Taktiež navrhujeme pridať dáta ako je analýza sentimentu a externé finančné ukazovatele, aby sa zlepšila zrozumiteľnosť dát a ich následná predikcia.
- Vnútorne parametre modelov:** Keďže neboli uvedené parametre ako napríklad veľkosť kernel-u alebo sekvencia pre časové rady, zvolili sme ich na základe viacerých testov štúdiou navrhutej validácie ale aj krížovej validácie.

## Vyhodnotenie výsledkov

- Naša replikácia poukázala na nezrovnalosti výsledkov na daných trhoch ako za použitia pôvodného prístupu, tak aj toho upraveného a tým vyvolala pochybnosť o reálnosti výsledkov implementácie metód štúdie aj keď princíp danej metodiky dáva zmysel.
- Už pri prvom pohľade na vstupné dáta sme si všimli ich nevyvážené rozloženie, ktoré sa nezmenilo ani po pridaní technických ukazovateľov a aplikácii PCA, kde sme pozorovali iné správanie kriviek v trénovacej množine dát ako u testovacej.
- Inicializácia váh bola dôležitým faktorom, kde niektorá spôsobovala záporný nárast ceny, a teda sa trénovaný model reštartoval, v prípadoch kedy nastal takýto problém.

<i>PCA – LSTM – CNN</i>	<i>study MSE</i>	<i>our MSE</i>	<i>MAPE</i>	<i>MAE</i>	<i>R<sup>2</sup></i>
30 dňový posun	0.0011	0.0473	492.184	0.1681	0.5140
60 dňový posun	0.0028	0.0493	$2941 \cdot 10^{11}$	0.1603	0.4902
Křížová validácia 30 dní	-	0.1876	$5773 \cdot 10^{10}$	0.3710	-3.8011
Křížová validácia 60 dní	-	0.1524	$6261 \cdot 10^{10}$	0.3361	-2.8335



Graf 1: Vizualizácia výsledkov