

**M A S A R Y K O V A
U N I V E R Z I T A**

FAKULTA INFORMATIKY

**Analytická platforma pre
spracovanie finančných správ
a predikciu správania akciového
trhu pomocou metód strojového
učenia**

Diplomová práca

SAMUEL ŠULO

Brno, jar 2026

**M A S A R Y K O V A
U N I V E R Z I T A**

FAKULTA INFORMATIKY

**Analytická platforma pre
spracovanie finančných správ
a predikciu správania akciového
trhu pomocou metód strojového
učenia**

Diplomová práca

SAMUEL ŠULO

Vedúci práce: doc. RNDr. Aleš Horák, Ph.D.

Katedra strojového učenia a spracovania dát

Brno, jar 2026



Vyhlásenie

Vyhlasujem, že táto diplomová práca je mojím pôvodným autorským dielom, ktoré som vypracoval samostatne. Všetky zdroje, pramene a literatúru, ktoré som pri vypracovaní používal alebo z nich čerpal, v práci riadne citujem s uvedením úplného odkazu na príslušný zdroj. Pri písaní zdrojového kódu a niektorých textových častí práce som využíval nástroj GitHub Copilot ako asistenciu umelej inteligencie. Po použití tohto nástroja som vykonal kontrolu obsahu a preberám zaň plnú zodpovednosť.

Samuel Šulo

Vedúci práce: doc. RNDr. Aleš Horák, Ph.D.

Podakovanie

Rád by som sa poďakoval doc. RNDr. Alešovi Horákovi, Ph.D. za odborné vedenie a cenné rady pri tvorbe tejto diplomovej práce.

Zhrnutie

Práca sa zaoberá návrhom a implementáciou analytickej platformy FinPulse zameranej na spracovanie finančných správ a predikciu správania akciového trhu pomocou metód strojového učenia. V jej rámci sú riešené dve predikčné úlohy — sentimentová klasifikácia finančných textov a predikcia smeru pohybu ceny na nasledujúci obchodný deň. Pre sentimentovú klasifikáciu sú realizované porovnania logistickej regresie a LSTM s predtrénovaným FinBERTom ako referenčným modelom zodpovedajúcim súčasnému štandardu spracovania finančného textu. Pre predikciu pohybu ceny sú porovnávané logistická regresia, LightGBM a LSTM. Výstupy modelov sú integrované do webovej aplikácie, ktorá ich spája s externými dátovými zdrojmi a zobrazuje v analytickom rozhraní FinPulse. Modely sú vyhodnotené na verejne dostupných finančných dátach pomocou štandardných klasifikačných metrík.

Kľúčové slová

strojové učenie, spracovanie prirodzeného jazyka, sentimentová analýza, predikcia pohybu ceny akcií, finančné správy, Financial PhraseBank, FinBERT, LSTM, LightGBM, analytická platforma

Obsah

Úvod	1
1 Teoretické základy analýzy finančných trhov	3
1.1 Princípy fungovania a analýzy akciových trhov	3
1.2 Behaviorálne financie a vplyv sentimentu	6
1.3 Evolúcia spracovania prirodzeného jazyka vo financiách	10
1.4 Strojové učenie v predikcii finančných časových radov .	12
2 Návrh platformy FinPulse	16
2.1 Analýza požiadaviek a špecifikácia funkcionality	16
2.2 Celková architektúra platformy	17
2.3 Dátové zdroje a integrácia externých API	19
2.3.1 Trhové a spravodajské dáta	20
2.3.2 Jazykové modely a sémantická analýza	20
2.3.3 Portfóliové a fundamentálne dáta	21
2.4 Návrh ML pipeline	22
2.4.1 Sentimentová vetva	22
2.4.2 Predikčná vetva a evaluácia	23
2.4.3 Cloudová infraštruktúra a nasadenie	23
2.5 Návrh používateľského rozhrania	24
3 Implementácia	25
3.1 Dátová príprava a predspracovanie	25
3.2 Architektúra a tréovanie modelov analýzy sentimentu	27
3.3 Konštrukcia price-move modelov a fúzia dát	29
3.4 Serverová logika a konektivita externých služieb	31
3.5 Klientska aplikácia a vizualizácia dát	35
4 Experimenty a vyhodnotenie	37
4.1 Datasety a experimentálny protokol	37
4.2 Vyhodnotenie sentimentových modelov	38
4.2.1 Kvantitatívne výsledky	38
4.2.2 Analýza chýb	39
4.2.3 Vzťah k modelu FinBERT	40
4.2.4 Kvalitatívna inferencia	40
4.2.5 Zhrnutie	41

4.3	Vyhodnotenie modelov predikcie cenového pohybu . . .	42
4.3.1	Kvantitatívne výsledky	43
4.3.2	Jednoakciový vs. združený tréning	43
4.3.3	Vplyv sentimentových príznakov	44
4.3.4	Dôležitosť príznakov (LightGBM)	45
4.3.5	Diskusia	46
4.3.6	Zhrnutie	47
4.4	Praktická relevantnosť a limity modelov	48
4.4.1	Čo z metrík plyní pre rozhodovanie	48
4.4.2	Limity dát a výberu vzorky	48
4.4.3	Možné skreslenia	49
4.4.4	Interpretácia výstupov v kontexte aplikácie . . .	49
Záver		51
Bibliografia		53
A Snímky používateľského rozhrania		57
B Prompty pre OpenAI API		61
B.1	Filter relevantnosti správy	61
B.2	Analýza článku	61
B.3	Investičný súhrn aktíva	63
B.4	Analýza portfólia	64

Zoznam tabuliek

4.1	Porovnanie sentimentových modelov na validačnej a testovacej množine.	38
4.2	Metriky po triedách na testovacej množine.	39
4.3	Inferencia sentimentu na troch vzorových textoch.	40
4.4	Porovnanie modelov predikcie cenového pohybu.	43
4.5	Desať príznakov podľa počtu rozdelení v stromoch.	46

Zoznam obrázkov

2.1	Celková architektúra platformy	19
A.1	Prehľadová obrazovka <i>Overview</i> (dashboard, správy, ukazovatele).	57
A.2	Obrazovka <i>Portfolio</i> po synchronizácii s brokerom.	58
A.3	Obrazovka <i>Asset Tracker</i> (sledované tituly).	58
A.4	Obrazovka <i>Analyst Targets</i> (index cieľových cien).	59
A.5	Obrazovka <i>Calendar</i> (výsledky, dividendy, makro).	59
A.6	Obrazovka <i>Market Caps</i>	60
A.7	Detail konkrétneho titulu.	60

Úvod

Množstvo informácií, ktoré každý deň prechádzajú globálnymi finančnými trhmi, dávno prekročilo hranicu toho, čo dokáže jednotlivý investor spracovať v reálnom čase. Stovky tisíc správ, výkazov a analytických reportov ovplyvňujú ceny aktív, niekedy v priebehu sekúnd. Tradičné prístupy k investičnému rozhodovaniu, opierajúce sa o manuálne sledovanie správ a základnú technickú analýzu, čoraz viac narážajú na tento informačný strop. Práve tu vzniká priestor pre automatizované analytické systémy schopné spracovať veľké objemy neštruktúrovaných dát a premeniť ich na interpretovateľné predikcie. Pri analýze finančných textov však pojem sentimentu neznamena iba emočný odtieň bežnej komunikácie z recenzií či sociálnych sietí, ale aj orientáciu vyhlásení voči príslušnej spoločnosti a jej očakávanému dopadu z pohľadu investorov pri zachovaní významu účtovných a trhových pojmov, ktoré mimo finančnej domény vyvolávajú odlišnú polaritu (*finančný sentiment*). Rovnaký výraz môže byť vo finančnom kontexte neutrálny, hoci ho model založený na všeobecnom slovníku vyhodnotí ako negatívny.

Motiváciou tejto práce je skutočnosť, že existujúce riešenia stoja pred zásadnou voľbou: buď sú metodologicky hodnotné, no neprístupné bežnému používateľovi, alebo sú dostupné, no metodologicky obmedzené. Profesionálne platformy ako Bloomberg Terminal či Refinitiv poskytujú pokročilú analytiku vrátane sentimentových indikátorov, avšak ich licenčné náklady ich robia nedostupnými pre individuálnych investorov. Voľne dostupné nástroje, ako FinViz alebo TradingView, naopak ponúkajú prehľadné rozhranie, no bez hlbšej integrácie modelov strojového učenia a automatizovanej analýzy správ. FinPulse rieši oba nedostatky spojením výskumne relevantného modulu strojového učenia s interaktívnou webovou aplikáciou.

Cieľom práce je navrhnúť a implementovať analytickú platformu FinPulse, ktorá spracúva finančné správy a predikuje cenové pohyby akcií. FinPulse pozostáva z dvoch vzájomne prepojených vrstiev. Prvou je modul strojového učenia riešiaci dve predikčné úlohy — kategorizáciu sentimentu finančných textov a predikciu smeru pohybu ceny (UP/DOWN) na nasledujúci obchodný deň. Pre sentimentovú klasifikáciu sú implementované a porovnávané logistická regresia a LSTM.

FinBERT je zaradený ako predtrénovaný referenčný model zodpovedajúci súčasnému štandardu spracovania finančného textu [1]. Pre predikciu cenového pohybu sú porovnávané logistická regresia, LightGBM a LSTM. Druhou vrstvou je webová aplikácia, ktorá integruje výstupy modulu strojového učenia s externými dátovými zdrojmi a poskytuje jednotné používateľské rozhranie k analytickým modulom FinPulse.

Práca je štruktúrovaná nasledovne. Kapitola 1 buduje teoretický základ v oblasti spracovania prirodzeného jazyka vo finančnej doméne a algoritmických prístupov strojového učenia. Kapitola 2 sa venuje analýze požiadaviek a plánovanej funkcionality, návrhu platformy FinPulse a špecifikácii jej komponentov. Kapitola 3 dokumentuje implementáciu modulu strojového učenia, serverovej logiky s napojením na externé služby a klientskej aplikácie. Kapitola 4 prezentuje experimentálne výsledky a ich interpretáciu.

1 Teoretické základy analýzy finančných trhov

Automatizovaná analýza finančných trhov sa v posledných dvoch desaťročiach posunula od jednoduchých pravidlových systémov k sofistikovaným modelom strojového učenia, ktoré dokážu spracúvať heterogénne zdroje dát vrátane neštruktúrovaného textu finančných správ. Rast objemu a rýchlosti publikovania finančných informácií zároveň vytvoril prostredie, v ktorom manuálna analýza prestáva byť škálovateľná a algoritmické prístupy získavajú praktický aj akademický význam [2].

Kapitola je členená do štyroch tematických celkov. Prvý sa venuje princípom fungovania akciových trhov — hypotéze efektívnych trhov, technickej a fundamentálnej analýze — a vymedzuje teoretický priestor, v ktorom môžu mať cenové a textové signály prediktívnu hodnotu [3]. Druhý celok sa zameriava na behaviorálne financie a sentiment: vysvetľuje, prečo sú iracionality investorov systematické a merateľné, rozlišuje medzi individuálnym, trhovým a textovým sentimentom a predstavuje rozšírené dimenzie analýzy správ — typ udalosti, zmenu pozornosti trhu, abnormalitu sentimentu voči historickému priemeru a časový horizont dopadu. Tretí celok sleduje evolúciu spracovania prirodzeného jazyka vo finančnej doméne od metód Bag-of-Words cez slovné vnorenia až po transformerové modely a ich doménovo adaptované varianty [1]. Štvrtý celok sa venuje strojovému učeniu v predikcii finančných časových radov: charakterizuje výzvy spojené s nelinearitou, nestacionaritou a šumom, opisuje rôznorodosť vstupných dátových zdrojov a predstavuje rekurentné neurónové siete a architektúru LSTM ako hlavný prístup pre sekvenčné finančné predikcie [4].

1.1 Princípy fungovania a analýzy akciových trhov

Akciové trhy plnia v ekonomike dvojakú úlohu: umožňujú spoločnostiam získavať kapitál a zároveň investorom participovať na ich raste. Ceny akcií na týchto trhoch nie sú určované jedným faktorom, ale vznikajú ako výsledok neustálej interakcie medzi ponukou a dopytom, informačnými tokmi, makroekonomickými podmienkami a správaním samotných účastníkov trhu. Práve táto komplexnosť formácie

cien vyústila do niekoľkých teoretických rámcov, ktoré sa pokúšajú vysvetliť, do akej miery sú pohyby cien predvídateľné a aké analytické metódy je vhodné pri ich skúmaní používať.

Hypotéza efektívnych trhov a teória náhodnej prechádzky Základným teoretickým kameňom modernej finančnej ekonómie je Hypotéza efektívnych trhov (EMH), ktorú Eugene Fama v roku 1970 systematicky sformuloval v prehľadovej práci, kde spája teoretický rámec so syntézou dostupných empirických štúdií. EMH vychádza z predpokladu, že finančné trhy sú informačne efektívne — ceny cenných papierov v každom okamihu odrážajú všetky dostupné informácie, takže žiadny investor nemôže dlhodobo a systematicky dosahovať nadpriemerné výnosy. Fama rozlíšil tri formy trhovej efektívnosti: slabú, polosilnú a silnú. Slabá forma predpokladá, že historické ceny neobsahujú žiadnu prediktívnu hodnotu pre budúce pohyby, čo je konzistentné s teóriou náhodnej prechádzky — trhové zmeny sú vzájomne nezávislé a nepredvídateľné. Polosilná forma rozširuje tento predpoklad na všetky verejne dostupné informácie vrátane finančných správ, výsledkov hospodárenia a makroekonomických oznamov. Silná forma, ktorú Fama sám považoval skôr za okrajový teoretický scenár bez bežnej empirickej platnosti, tvrdí, že ceny absorbujú dokonca aj súkromné informácie [5].

Napriek tomu intelektuálna dominancia EMH začala slabnúť začiatkom 21. storočia. Rastúci počet empirických štúdií identifikoval trhové anomálie, ktoré s hypotézou efektívnych trhov nie sú konzistentné — napríklad efekt momentu, januárový efekt či systematický vplyv sentimentu investorov na ceny. Behaviorálna ekonómia tieto pozorovania vysvetlila prostredníctvom psychologických faktorov: investori nie vždy konajú racionálne, podliehajú kognitívnym skresleniam, stádovitému správaniu a emocionálnym reakciám na nové informácie [2]. Dnes preto prevláda konsenzus, že trhy nie sú ani dokonale efektívne, ani úplne chaotické — priestor pre nenáhodné predikcie existuje, no je obmedzený a nestabilný [3]. Práve na tieto medzery sa zameriavajú prístupy strojového učenia kombinujúce cenové a textové signály, ktorým sa venujú nasledujúce kapitoly.

Technická analýza Jedným z najrozšírenejších prístupov k predikcii trhových pohybov je technická analýza, ktorá vychádza z predpokladu, že historické cenové vzorce a obchodné objemy obsahujú prediktívnu informáciu o budúcom vývoji. Na rozdiel od EMH v jej slabej forme teda technická analýza predpokladá, že trhy vykazujú určitú mieru zotrvačnosti a opakujúcich sa štruktúr. Technickí analytici sa nesústreďia na fundamentálnu hodnotu spoločnosti, ale na samotný pohyb ceny a relevantné štatistické indikátory odvodené z historických dát [2].

Medzi najpoužívanjšie nástroje patria pohyblivé priemery (MA), ktoré filtrujú krátkodobý šum v cenových radoch a umožňujú identifikovať pretrvávajúce trendové smery. Skríženie dvoch MA s rozdielnou dĺžkou okna — napríklad 10-dňového a 30-dňového — sa klasicky používa ako signál zmeny trendu a patrí medzi základné obchodné pravidlá technickej analýzy [6]. Oscilátory ako Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Stochastický oscilátor a Bollingerove pásma zachytávajú momentum trhu, intenzitu trendu a mieru volatility [2]. RSI meria rýchlosť a veľkosť cenových pohybov a slúži na identifikáciu prekúpených alebo prepredaných stavov trhu, zatiaľ čo MACD zachytáva vzťah medzi krátkodobým a dlhodobým exponenciálnym pohyblivým priemerom. Tieto indikátory tvoria v kontexte tejto práce kľúčovú súčasť vstupných príznakov pre modely predikcie cenových pohybov a ich výpočet a normalizácia sú podrobne opísané v implementačnej časti.

Samotná technická analýza však nedokáže zachytiť externé informačné šoky, akými sú neočakávané správy alebo regulačné zmeny — a práve tu nastupuje fundamentálna analýza a modelovanie sentimentu z textových zdrojov.

Fundamentálna analýza Zatiaľ čo technická analýza pracuje výhradne s trhovými dátami, fundamentálna analýza hodnotí vnútornú ekonomickú hodnotu aktíva na základe širšieho súboru informácií. Patrí sem analýza finančných výkazov (výkazy ziskov a strát, súvahy, výkazy peňažných tokov), hodnotenie makroekonomických ukazovateľov — ako sú úrokové sadzby, inflácia, miera nezamestnanosti alebo rast HDP — a analýza správ a udalostí ovplyvňujúcich daný sektor alebo spoločnosť [2]. Kľúčovými metrikami sú trhovú kapitalizácia,

zisk na akciu (EPS), pomer ceny k zisku (P/E) alebo pomer dlhu k vlastnému kapitálu (D/E).

Zásadnou črtou fundamentálnej analýzy je jej orientácia na neštruktúrované textové dáta. Správy, výročné správy, prepisy zárobných konferencií a analytické reporty predstavujú bohatý informačný zdroj, ktorého systematické spracovanie je nad ľudské kapacity. Výskum opakovane potvrdil, že existuje tesná korelácia medzi pohybmi cien akcií a správami súvisiacimi s príslušnými spoločnosťami [4], pričom modely kombinujúce numerické trhové dáta s textovými príznakmi systematicky prekonávajú prístupy pracujúce len s jedným typom vstupu. Dôvodom je, že historický trend zachytáva momentálne trhové impulzy, zatiaľ čo sentiment správ zachytáva informačné šoky, ktoré sa ešte nestačili premietnuť do historických cien.

Technická a fundamentálna analýza nie sú v praxi konkurenčné, ale komplementárne perspektívy: prvá zachytáva kvantitatívne impulzy trhu, druhá hodnotí kvalitatívny informačný kontext. Ich prepojenie spočíva v kombinovaní cenových a textových vstupov a tvorí metodologickú os prístupu, ktorý je v práci ďalej rozvíjaný.

1.2 Behaviorálne financie a vplyv sentimentu

Klasická finančná ekonómia dlho vychádzala z predpokladu, že investori sú racionálni aktéri, ktorí pri rozhodovaní maximalizujú očakávanú užitočnosť na základe všetkých dostupných informácií. Tento predpoklad je základným pilierom Hypotézy efektívnych trhov, avšak empirická skúsenosť ho opakovane spochybňuje. Behaviorálna ekonómia ponúka alternatívny pohľad: investori sú ľudia, nie algoritmy, a ich rozhodnutia sú formované nielen faktami, ale aj emóciami, kognitívnymi skresleniami a sociálnymi vplyvmi. Dnes je konsenzus v literatúre jasný — debata o prediktabilitě trhov sa presunula od otázky, či sentiment investorov trhy ovplyvňuje, k otázke, ako ho merať a kvantifikovať jeho efekt [3].

Psychológia investora a iracionálne správanie Behaviorálna ekonómia identifikuje celý rad systematických odchýlok od racionálneho správania, ktoré vedú k predvídateľným trhovým anomáliám. Medzi najvýznamnejšie patria efekt zakotvenia (*anchoring*), pri ktorom sa

investori príliš spoliehajú na prvú dostupnú informáciu pri tvorbe úsudku, efekt stáda (*herding*), keď jednotlivci nasledujú správanie davu aj napriek vlastnému skepticizmu, a konfirmačná predpojatosť, pri ktorej investori selektívne vyhľadávajú informácie potvrdzujúce ich existujúce presvedčenia [2]. Tieto mechanizmy majú priamy dopad na dynamiku cien a volatilitu trhu. Napríklad stádové správanie môže zosilňovať cenové pohyby nad rámec fundamentálnych hodnôt, čím vznikajú bubliny alebo panické výpredaje, ktoré nie sú ospravedlňiteľné čisto ekonomickými faktormi. Behaviorálna finančná teória navyše argumentuje, že rozhodnutia investorov nie sú formované len samotným obsahom správ, ale aj spôsobom ich podania, kontextom a emocionálnym ladením — čo otvára priestor pre kvantitatívnu analýzu sentimentu ako prediktívneho nástroja [7].

Z pohľadu predikcie akciových trhov je dôležité, že tieto iracionality sú systematické, nie náhodné. Systematické skreslenia vytvárajú vzory, ktoré možno identifikovať a modelovať. Práve toto je teoretickým základom pre využitie algoritmov strojového učenia pri predikcii trhových pohybov — nie ako nástroja na využívanie individuálnych chýb, ale ako prostriedku na zachytenie agregatívneho sentimentu, ktorý sa premieta do cien aktív [2].

Individuálny, trhový a textový sentiment V literatúre sa rozlišujú tri vzájomne previazané, avšak konceptuálne odlišné pojmy: individuálny sentiment investora, trhový sentiment a textový finančný sentiment [3]. Individuálny sentiment investora vyjadruje jeho optimizmus alebo pesimizmus ohľadom budúcej trhovej aktivity a môže byť meraný prostredníctvom prieskumov nálad medzi investormi alebo odvodený z finančných textov. Trhový sentiment je kolektívnym odrazom investorských nálad a prejavuje sa v obchodnom správaní. Za *býči trh* sa v odbornej aj laickej terminológii označuje dlhšie obdobie, keď väčšina účastníkov očakáva ďalší rast kurzov alebo indexov a nakupuje v presvedčení o budúcom zisku z predaja, zatiaľ čo pojem *medvedí trh* sa viaže na obdobia s prevahou pesimistických očakávaní výraznejšieho poklesu cien a výpredajov. V praxi sa tieto označenia používajú na zachytenie súhrnného smeru vývoja širších trhov alebo indexov, nie na opis izolovaných jednodňových pohybov. Merateľný

je prostredníctvom ukazovateľov ako CBOE Volatility Index (VIX), Baker-Wurgler Index alebo Put-Call Ratio.

Textový finančný sentiment je meraný mierou pozitivity alebo negativity v obsahu finančných textov. Nie je synonymom trhového sentimentu, ale medzi oboma existuje tesné prepojenie: finančný textový sentiment obsahuje subjektívne hodnotenia investorov a analytikov, ako aj objektívne informácie vo forme správ a výkazov. Objektívne informácie, napríklad politické a makroekonomické správy, prelomové udalosti a výročné správy vydávané spoločnosťami, odrážajú podmienky vo všeobecnom prostredí, odvetviach, trhoch a firmách. Oproti samotnému trhovému sentimentu majú typicky časový predstih v tom zmysle, že formujú úsudky investorov ešte pred tým, ako sa premietnu do cien [3]. Subjektívne hodnotenia sú naopak oneskorené a odrážajú nálady, ktoré trh do istej miery už spracoval. Tento časový rozmer má praktické implikácie pre konštrukciu prediktívnych modelov: správne zarovnanie textového sentimentu k príslušnému obchodnému dňu je nevyhnutnou podmienkou validnosti predikcie.

Dôležitým poznatkom je aj asymetrický charakter vplyvu sentimentu na trhy. Výskum opakovane dokumentuje, že negatívne správy majú spravidla väčší a dlhší dosah na ceny aktív ako správy pozitívne. Zatiaľ čo pozitívne informácie sú typicky asociované s krátkodobým skokovým nárastom ceny, efekt negatívnych správ sa rozkladá do dlhšieho časového horizontu [3]. Tento asymetrický vzorec má priame dôsledky pre výber a hodnotenie modelov sentimentovej analýzy: model, ktorý systematicky zlyháva pri identifikácii negatívneho sentimentu, podceňuje práve ten typ signálu, ktorý má na trhovú dynamiku najväčší dopad.

Rozšírené dimenzie textového sentimentu Samotná polarita sentimentu — pozitívny, negatívny, neutrálny — predstavuje len základnú vrstvu informácie obsiahnutej v správach. Prediktívnu hodnotu možno zvýšiť zapojením ďalších dimenzií. Typ udalosti, teda či ide o zverejnenie výsledkov hospodárenia, regulačný zásah, akvizíciu alebo makroekonomické oznámenie, moduluje silu a smer trhovej reakcie, pretože každá kategória udalosti nesie odlišný informačný obsah a vyvoláva odlišnú intenzitu trhovej pozornosti [3]. Zmena pozornosti trhu, meraná napríklad náhlým nárastom objemu správ

o danom titule, môže predchádzať samotnej cenovej reakcii a slúžiť ako včasný signál zvýšenej neistoty [7]. Abnormalita sentimentu voči historickému priemeru — odchýlka od bežnej hladiny pozitivity alebo negativity pre daný titul — signalizuje informačný šok, ktorý trh pravdepodobne ešte plne neabsorboval, a preto môže mať vyššiu prediktívnu hodnotu ako absolútna hodnota sentimentu samotného. Časový dopad správ napokon určuje, v akom horizonte sa efekt správy prejaví na cene: krátkodobé udalosti, akými sú prekvapivé výsledky hospodárenia, vyvolávajú okamžitú reakciu, zatiaľ čo makroekonomické zmeny pôsobia na dlhšom horizonte. Tieto dimenzie spoločne formujú rámec pre konštrukciu obohateného príznakového priestoru v prediktívnych modeloch [2].

Korelácia správ a trhových pohybov Vzťah medzi zverejnením informácií a okamžitou reakciou trhu je jedným z najdôkladnejšie skúmaných javov vo finančnom výskume. Výsledky opakovane potvrdzujú existenciu tesnej korelácie medzi správami súvisiacimi s príslušnými spoločnosťami a pohybmi ich akciových cien [4]. Typickými príkladmi sú zverejnenia výsledkov hospodárenia, zmeny v manažmente alebo akvizície, ktoré často podliehajú požiadavke na bezodkladné informovanie trhu a majú okamžitý a merateľný dopad na trhovú hodnotu. Práve tento dopad je teoretickým zdôvodnením, prečo sentiment správ môže slúžiť ako prediktívny signál.

Experimentálne dôkazy idú ešte ďalej. Modely kombinujúce historické cenové dáta so sentimentovými príznakmi zo správ môžu pri dostatočnom objeme a kvalite textových dát prekonávať prístupy pracujúce výlučne s numerickými vstupmi [4]. Dôvodom je, že historické ceny zachytávajú impulzy, ktoré trh už absorboval, zatiaľ čo textový sentiment správ zachytáva informačné šoky, ktoré sa do cien ešte nestihli premietnuť. Tento mechanizmus je obzvlášť výrazný v prípade neočakávaných udalostí — regulačných zmien, korporátnych škanďálov alebo makroekonomických prekvapení — kde reakcia trhu na správu predchádza akémukoľvek historickému cenovému signálu.

Tieto zistenia priamo motivujú integráciu dvoch komplementárnych zdrojov informácií v analytických systémoch tohto typu: histórie cien a objemov obchodov na jednej strane a sentimentového skóre finančných správ na strane druhej. Teoretický základ pre túto kom-

bináciu je v literatúre opodstatnený. Experimentálne overenie toho, do akej miery konkrétne algoritmy tento vzťah v praxi zachytia, je predmetom kapitoly 4.

1.3 Evolúcia spracovania prirodzeného jazyka vo financiách

Schopnosť automaticky extrahovať informácie z neštruktúrovaného textu tvorí technologický základ modulu strojového učenia v analytických platformách tohto typu, akou je v tejto práci FinPulse. Spracovanie prirodzeného jazyka vo finančnej doméne pritom prešlo za posledné dve desaťročia výraznou evolúciou — od jednoduchých štatistických metód cez distribučné reprezentácie slov až po kontextualizované transformerové modely, ktoré v súčasnosti predstavujú štandard odvetvia.

Tradičné štatistické metódy Prvá generácia NLP prístupov vo financiách vychádzala z predpokladu, že dokument možno opísať ako množinu slov bez ohľadu na ich poradie. Metóda Bag-of-Words reprezentuje text ako vektor frekvencií výskytu jednotlivých slov v korpuse, pričom gramatická štruktúra ani sémantický kontext sa pri tomto prístupe nezohľadňujú [8]. TF-IDF (*Term Frequency — Inverse Document Frequency*) tento prístup rozširuje o váhovanie slov podľa ich špecificity v korpuse — slová vyskytujúce sa v mnohých dokumentoch získavajú nižšiu váhu, zatiaľ čo výrazy typické pre konkrétny text sú zvýraznené [9, 8]. Tieto metódy majú zásadnú výhodu: sú výpočtovo nenáročné a dobre interpretovateľné. Ich použitie pretrváva dodnes najmä ako rýchly baseline pri porovnaní modelov, čo potvrdzujú aj štúdie, v ktorých TF-IDF s logistickou regresiou dosahuje konkurencieschopné výsledky pri jednoduchých klasifikačných úlohách [2].

Vo finančnej doméne však tieto metódy narážajú na zásadné obmedzenia. Finančný text sa vyznačuje vysokou hustotou odbornej terminológie, stručnosťou a smerovou závislosťou výrazov — rovnaké slovo môže v závislosti od kontextu vyjadrovať odlišnú polaritu sentimentu [3]. Bag-of-Words a TF-IDF nedokážu zachytiť tieto nuansy, pretože slová posudzujú izolovane. Typickým príkladom je negácia vo vete „žiadne straty neboli zaznamenané“, kde sa polarita sentimentu

určuje až z celej vety spolu s negáciou, nie z izolovaného výrazu o strastách. Ak sa text skráti alebo zostanú len vytrhnuté slová bez zachovanej syntaxe a negácie, odhad sentimentu sa ľahko obráti. Preto je v tejto doméne nevyhnutné prejsť k modelom zachytávajúcim kontext [1].

Slovné vnorenia Zásadný posun prinieslo zavedenie neurónových vektorových reprezentácií slov, označovaných ako *word embeddings*. Modely Word2Vec [10] a GloVe (*Global Vectors for Word Representation*) [11] reprezentujú každé slovo ako hustý vektor v nízkorozmernom priestore, pričom sémanticky príbuzné slová majú navzájom bližšie vektory. Na rozdiel od *one-hot* kódovania, pri ktorom každé slovo zodpovedá jednému rozmeru vektora, *word embeddings* produkujú vektory s niekoľkými stovkami dimenzií — tým výrazne znižujú počet parametrov, ktoré musia algoritmy strojového učenia odhadnúť.

Kľúčová vlastnosť týchto reprezentácií spočíva v tom, že zachytávajú sémantické vzťahy: slová, ktoré sa objavujú v podobných kontextoch, sú priradené podobným vektorom. V prostredí Word2Vec by napríklad slová „akcia“ a „dividenda“ boli umiestnené bližšie k sebe ako k výrazu „automobil“. Pre finančné NLP to znamenalo výrazné zlepšenie pri zachytávaní doménovej terminológie. Nevýhodou ostáva, že každé slovo je reprezentované jediným statickým vektorom bez ohľadu na okolitý kontext — slovo „zámok“ má rovnaký vektor bez ohľadu na to, či ide o šľachtické sídlo alebo o zamykací mechanizmus na dverách.

Transformery a veľké jazykové modely Definitívnu zmenu paradigmy priniesla architektúra Transformer [12]. Jej jadro tvorí mechanizmus pozornosti (*attention mechanism*), ktorý modelu umožňuje pri generovaní reprezentácie každého slova dynamicky zvažovať vzťahy ku všetkým ostatným slovám vo vete — čím vznikajú kontextualizované *embeddings*. Na rozdiel od statických vektorov *word2vec* tak model priradzuje rovnakému slovu odlišnú reprezentáciu v závislosti od celého okolitého textu. Model BERT (*Bidirectional Encoder Representations from Transformers*) od Googlu [13] tento princíp rozšíril o obojsmerné spracovanie kontextu — pri kódovaní každého tokenu berie do úvahy slová pred ním aj za ním súčasne.

Pre finančnú doménu je však rozhodujúci posun od všeobecného BERT modelu k doménovo adaptovaným variantom. Všeobecný BERT je predtrénovaný na Wikipédii a korpuse kníh, čo znamená, že jeho vektorový priestor odráža všeobecný jazyk — nie odborný finančný diskurz. FinBERT [14] rieši tento problém predtréningom na rozsiahlom korpuse finančných textov zahŕňajúcom výročné správy, analytické reporty a prepisy zárodkových konferencií. Výsledkom je model, ktorý rozumie špecifickému lexikónu a štýlu finančného písania a dokáže rozlíšiť, že výraz „allowance“ má v súvahe oproti bežnej komunikácii odlišnú sentimentovú polaritu. Práve táto doménová adaptácia je hlavným dôvodom, prečo FinBERT dosahuje na testovacej sade 1 000 manuálne anotovaných viet z analytických správ presnosť sentimentovej klasifikácie 88,2 % v porovnaní s 62,1 % slovníkovej metódy (Loughran–McDonald) a 76,3 % LSTM [15, 1].

Rodina GPT modelov (*Generative Pre-trained Transformer*) rozširuje tento prístup o generatívnu schopnosť — model nielen klasifikuje, ale dokáže aj tvoriť súvislý text [16]. V analytických systémoch tohto typu GPT modely plnia úlohu generatívnej vrstvy zodpovednej za syntetizovanie súhrnov a formulovanie štruktúrovaných odporúčaní, čím dopĺňajú klasifikačné modely o ľudsky čitateľný výstup.

Evolúcia od Bag-of-Words k transformerom teda nepredstavuje lineárne zlepšenie jednej metriky, ale kvalitatívnu zmenu v schopnosti modelu porozumieť jazyku — od štatistiky výskytu slov až po pochopenie kontextu, sémantiky a doménových nuáns. Každá úroveň tejto hierarchie nachádza uplatnenie v závislosti od dostupných dát, výpočtových zdrojov a požadovanej interpretovateľnosti, čo robí voľbu reprezentácie kľúčovým rozhodnutím pri návrhu systémov pre finančnú textovú analýzu.

1.4 Strojové učenie v predikcii finančných časových radov

Finančné časové rady patria medzi najnáročnejšie dátové štruktúry, s akými sa metódy strojového učenia stretávajú. Na rozdiel od mnohých iných aplikačných domén, kde sú vstupné dáta relatívne stabilné a štruktúrované, cenové rady akcií vykazujú niekoľko vzájomne sa osilňujúcich vlastností, ktoré predikciu systematicky sťažujú. Pochope-

nie týchto vlastností je nevyhnutným predpokladom pre zodpovedný výber algoritmov a správnu interpretáciu ich výstupov.

Charakteristika finančných dát Prvou a najzásadnejšou výzvou je nelinearita vzťahov medzi vstupnými premennými a budúcou cenou aktíva. Tradičné lineárne modely ako ARIMA predpokladajú, že budúce hodnoty sú lineárnou funkciou hodnôt minulých — predpoklad, ktorý finančné trhy systematicky porušujú [6]. Ceny akcií sú výsledkom interakcie tisícov agentov s rôznymi informáciami, horizontmi a motiváciami, čo generuje nelineárne dynamiky, ktoré lineárne modely nedokážu zachytiť. Preto výskum opakovane dokumentuje, že hlboké neurónové siete schopné modelovať nelineárne vzťahy prekonávajú klasické ekonometrické prístupy v úlohách predikcie smeru cenového pohybu [2].

Druhou kritickou vlastnosťou je nestacionarita. Stacionárny časový rad má konštantnú strednú hodnotu, rozptyl a autokorelačnú štruktúru — vlastnosti, ktoré cenové rady akcií nemajú. Štatistické vlastnosti finančných radov sa menia v čase v závislosti od makroekonomického cyklu, regulačných zmien a trhových režimov [6]. Toto je praktickým dôvodom, prečo modely trénované na dátach z jedného trhového obdobia môžu výrazne zlyhať v období odlišnom — napríklad modely kalibrované v prostredí *býčieho trhu* zle predikujú správanie počas krízy. Metódy ako diferenciácia radov alebo normalizácia v kľzavých oknách túto nestacionaritu čiastočne adresujú, ale neodstraňujú ju úplne.

Treťou výzvou je vysoká miera šumu. Finančné časové rady obsahujú podstatne viac náhodných fluktuácií ako signálu. Pohyby cien sú z veľkej časti determinované novými informáciami, ktoré sú zo svojej podstaty nepredikovateľné — a práve táto nepredikovateľnosť je konzistentná s hypotézou efektívnych trhov v jej slabej forme [3]. Dôsledkom je, že horný strop presnosti klasifikácie smeru pohybu, ktorý možno od akéhokoľvek modelu reálne očakávať, je podstatne nižší než v iných aplikačných oblastiach hlbokého učenia. Hodnoty presnosti okolo 55–65 % sú v literatúre bežne reportovanými výsledkami a nie nevyhnutne znakom zlého modelu — môžu odrážať samotné limity predikovateľnosti [2].

Rôznorodosť a kvalita finančných dát Okrem technických výziev spojených s vlastnosťami cenových radov predstavujú finančné dáta problém aj z hľadiska ich rôznorodosti, spoľahlivosti a nerovnomernej informačnej hodnoty — tri pre túto doménu najrelevantnejšie dimenzie z rámca 5V, ktorým sa tradične charakterizujú veľmi veľké a heterogénne súbory dát (*Big Data*) [6]. Rôznorodosť sa prejavuje tým, že numerické cenové dáta, textové správy, sentimentové skóre, makroekonomické indikátory a technické indikátory tvoria štruktúrované aj neštruktúrované typy dát, ktoré je potrebné zlučovať do spoločného analytického rámca [4] — a práve preto jednoduché jednozdrojové modely systematicky zaostávajú za hybridnými prístupmi. Spoľahlivosť vstupov je ďalšou kritickou dimenziou: finančné správy môžu byť neúplné, chybné alebo zámerne skreslené, čo kladie mimoriadne nároky na predspracovanie dát. Napokon, hodnota jednotlivých dátových bodov nie je uniformná — nie každá správa je rovnako relevantná pre pohyb ceny konkrétnej akcie a rozlíšenie skutočného signálu od šumu zostáva jednou z kľúčových otvorených výziev [2].

Hlboké neurónové siete Odpoveďou výskumnej komunity na tieto výzvy bolo postupné nasadzovanie čoraz sofistikovanejších architektur hlbokého učenia, pričom rekurentné neurónové siete a ich rozšírenia sa stali dominantným prístupom. Základná architektúra RNN (*Recurrent Neural Network*) zavádza spätné väzby medzi skrytými stavmi siete, vďaka čomu môže pri spracovaní každého nového vstupu zohľadniť kontextové informácie z predchádzajúcich časových krokov [17]. Táto vlastnosť robí z RNN prirodzene vhodný nástroj pre sekvencné dáta vrátane finančných časových radov. V praxi sa vstupné sekvencie konštruujú metódou klzavého okna (*sliding window*): model pri každej predikcii vidí posledných n obchodných dní, pričom výskum ukazuje, že zohľadnenie historického kontextu systematicky zlepšuje výkon oproti prístupom pracujúcim len s jedným časovým krokom [4].

Tradičné RNN narážajú na fundamentálny problém, ktorý limituje ich praktickú využiteľnosť: problém miznúceho gradientu. Pri spätnom šírení chyby cez dlhé sekvencie sa gradienty exponenciálne zmenšujú, čo znemožňuje sieti naučiť sa dlhodobé závislosti [7]. Pre finančné aplikácie, kde môžu byť relevantné vzory vzdialené desiatky

obchodných dní od bodu predikcie, je toto obmedzenie závažné. Architektúra LSTM (*Long Short-Term Memory*) z roku 1997 rieši tento problém prostredníctvom mechanizmu brán [17]. Pamäťová bunka LSTM obsahuje vstupnú, zabúdaciú a výstupnú bránu, ktoré selektívne riadia zápis nových informácií do vnútorného stavu bunky, mieru zachovania jeho predchádzajúcej hodnoty a časť stavu, ktorá sa stane výstupom bunky. Táto štruktúra umožňuje sieti selektívne uchovávať alebo zabudnúť informácie podľa toho, čo je pre aktuálnu predikciu relevantné [7].

Empirické výsledky konzistentne potvrdzujú nadradenosť LSTM nad tradičnými prístupmi pri predikcii finančných časových radov. Systematický prehľad literatúry uvádza, že LSTM patrí medzi najčastejšie skúmané modely v tejto kategórii a vo väčšine komparatívnych štúdií dosahuje najlepšie výsledky spomedzi porovnávaných algoritmov [2]. Samostatné práce pritom ukazujú, že kombinácie LSTM s ďalšími modulmi architektúry alebo so sentimentovými príznakmi dokážu predikčné chyby ďalej znižovať oproti čistým sekvenčným či jednoduchým baseline modelom [7, 18].

LSTM tak zostáva dominantným prístupom pre sekvenčné finančné predikcie. V praktických systémoch je zvyčajne doplnený logistickou regresiou ako interpretovateľným baseline modelom a gradientovými ansámblovými metódami — napríklad LightGBM (*Light Gradient Boosting Machine*) — pre úlohy pracujúce so štruktúrovanými tabuľkovými príznakmi. Táto kombinácia pokrýva celé spektrum prístupov strojového učenia relevantných pre finančnú predikciu a umožňuje systematické porovnávanie modelov rôznych paradigiem.

2 Návrh platformy FinPulse

Finančné trhy každý deň produkujú obrovské množstvo informácií — správy, výsledky spoločností, komentáre analytikov, cenové pohyby. Spracovať všetky tieto vstupy manuálne je prakticky nemožné a práve tu vzniká priestor pre automatizované analytické systémy. Trh je zo svojej podstaty dynamický a nestabilný, pričom neočakávané zmeny okamžite ovplyvňujú ceny akcií — bez pokročilých nástrojov je teda spoľahlivé predpovedanie smeru pohybu len ťažko dosiahnuteľné [6].

Cieľom tejto kapitoly je popísať návrh platformy FinPulse, ktorá na tento problém reaguje. Návrh vychádza z potreby prepojiť neštruktúrované textové informácie z finančných správ s kvantitatívnymi trhovými údajmi a transformovať ich na interpretovateľné predikčné výstupy. Kapitola pokrýva analýzu požiadaviek a plánovanú funkcionality, celkovú architektúru platformy, integráciu externých dátových zdrojov, návrh ML pipeline vrátane cloudovej tréningovej infraštruktúry a koncepciu používateľského rozhrania.

2.1 Analýza požiadaviek a špecifikácia funkcionality

Pred návrhom architektúry bolo potrebné vymedziť, čo má platforma FinPulse používateľovi poskytovať a aké vstupy a výstupy pri tom predpokladá. Z literatúry a z cieľov práce vyplýva potreba jedného miesta, kde sa stretnú kvantitatívne trhové dáta, textové správy a výstupy modelov strojového učenia [4, 3]. Na tomto základe bola špecifikovaná nasledovná plánovaná funkcionality.

Spracovanie správ a sentiment. Systém má zbierať finančné správy z externých kanálov, ukladať ich a pre každú správu poskytovať automatickú analýzu: zhrnutie, odhad sentimentu, kategóriu udalosti a väzbu na dotknuté tituly alebo sektory. Okrem generatívnych a heuristických výstupov má byť k dispozícii aj doménovo špecializovaná klasifikácia sentimentu (FinBERT) a porovnateľné výstupy z vlastných natrénovaných modelov.

Predikcia smeru ceny. Pre vybrané tituly má FinPulse v rozhraní zobrazovať predikciu smeru pohybu ceny na ďalší obchodný deň, odvodenú z pipeline pracujúcej s cenovými a prípadne sentimentovými

príznakmi. Používateľ vidí výstup pripravený serverom, bez toho aby musel spúšťať tréning modelov.

Trhový prehľad a dátové rozšírenia. Plánované je zobrazenie aktuálnych trhových ukazovateľov, rebríčkov, ekonomického kalendára a ďalších kontextových modulov (cieľové ceny analytikov, trhové kapitalizácie), aby doplnili vlastné predikcie o verejne dostupné signály.

Personalizácia a účet. Používateľ sa má vedieť zaregistrovať a prihlásiť, spravovať profil, udržiavať zoznam sledovaných titulov (*watch-list*) a voliteľne prepojiť investičné portfólio cez *Trading212 API* [19], ktoré slúži na načítanie údajov o portfóliu do aplikácie vrátane prehľadu pozícií. Pri prehľade zloženia portfólia má FinPulse poskytovať AI podporu ako súčasť modulu portfólia v aplikácii.

Oddelenie tréningu a prevádzky. Požiadavka z výkonových a prevádzkových dôvodov: náročné trénovanie modelov má prebiehať mimo interaktívnej odozvy aplikácie (cloud), zatiaľ čo produkčná aplikácia len načítava artefakty a vykonáva inferenciu.

2.2 Celková architektúra platformy

Platforma FinPulse je navrhnutá ako trojvrstvová webová aplikácia s jasne oddelenými vrstvami pre perzistenciu dát, aplikačnú logiku a prezentáciu informácií. Toto rozdelenie nie je len štrukturálnou formalitou, ale nevyhnutnosťou pre zabezpečenie nezávislého vývoja a škálovateľnosti jednotlivých komponentov. Všeobecný rámec pipeline pre finančné aplikácie možno vnímať ako sekvenciu krokov od zberu dát cez predspracovanie až po evaluáciu, pričom návrh tohto toku priamo determinuje výslednú presnosť systému [18]. Predkladaný návrh tento rámec adaptuje na potreby FinPulse a rozširuje ho o vrstvu cloudovej trénovacej infraštruktúry.

Centrálnym orchestrátorom celého systému je asynchrónny backend implementovaný v jazyku Python [20] a postavený na frameworku FastAPI [21]. Voľba tohto nástroja bola podmienená potrebou spracovávať viacero paralelných požiadaviek s minimálnou latenciou. Backend musí súčasne obsluhovať dopyty z klientskej aplikácie, spúšťať inferenciu modelov a komunikovať s rozmanitými externými rozhraniami. Práve integrácia a správa dátových tokov z rôznych API predstavuje základný stavebný kameň systému, pričom podrobná

špecifikácia týchto zdrojov a spôsob ich napojenia sú rozobraté v podkapitole 2.3.

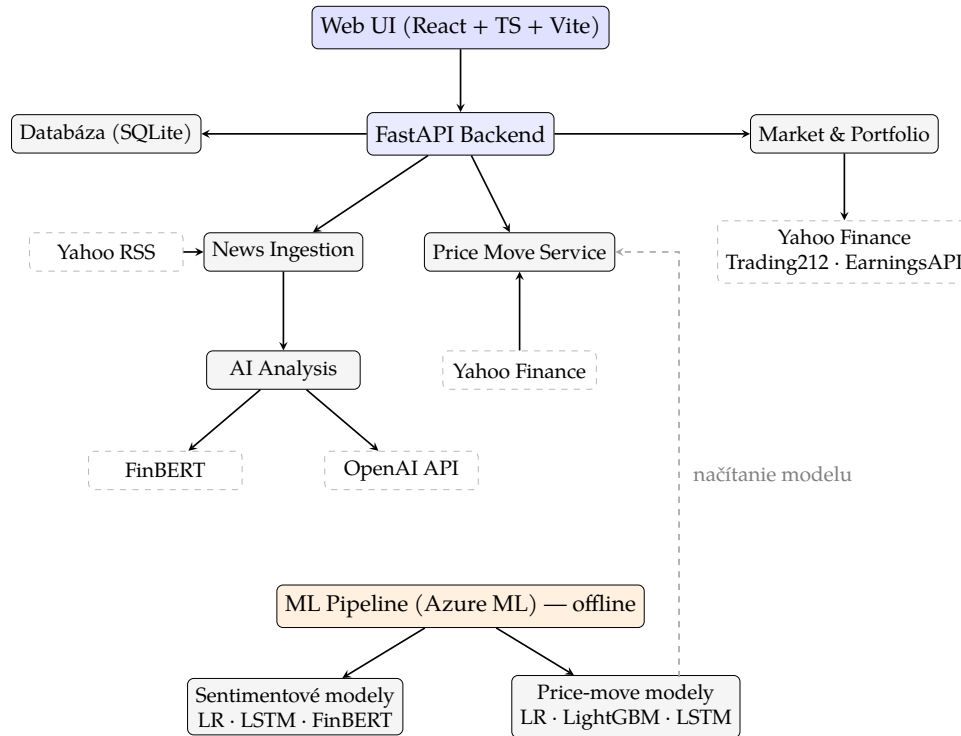
Dôležitým architektonickým rozhodnutím je oddelenie modulu strojového učenia od hlavného aplikačného jadra. Tento modul tvorí oddelenú sadu trénovacích skriptov a modelových artefaktov, zatiaľ čo produkčný backend tieto artefakty len načítava a zostáva responzívny pri obsluhu používateľských požiadaviek. Z pohľadu architektúry je kľúčové, že výpočtovo náročné operácie prebiehajú mimo hlavného aplikačného toku. Podrobný návrh životného cyklu modelov, cloudového tréningu a ML pipeline je uvedený v podkapitole 2.4.

Na úrovni perzistencie dát systém využíva databázu SQLite [22] prístupovanú cez vrstvu SQLAlchemy ORM [23]. Táto konfigurácia bola zvolená s ohľadom na jednoduchosť lokálneho nasadenia a možnosť budúcej migrácie na výkonnejšie relačné systémy. Databáza uchováva stav celej platformy FinPulse v jedenástich tabuľkách, zoskupených podľa domény nasledovne:

- **Účty a broker:**
users, trading212_connections
- **Sledované a portfóliové aktíva:**
watchlist, assets, portfolio_assets
- **Spravodajstvo a AI analýza:**
news_sources, articles, analyses
- **Súhrny aktív a portfólia:**
asset_daily_summaries, portfolio_analyses
- **Uložené predikcie pohybu ceny:**
price_move_predictions

Prezentačná vrstva je realizovaná v prostredí React [24] s využitím TypeScriptu [25]. Typová bezpečnosť je pri vývoji analytického ekosystému kľúčová, nakoľko eliminuje chyby pri spracovaní komplexných dátových štruktúr z modelov. Vizualizácia trhových trendov a interpretácia výstupov umelej inteligencie sú navrhnuté tak, aby investorovi poskytli okamžitý prehľad o dianí na trhu. Koncepcia tohto rozhrania a jednotlivé funkčné moduly sú predmetom podkapitoly 2.5.

Celkový tok informácií v rámci architektúry zachytáva diagram na obrázku 2.1.



Obr. 2.1: Celková architektúra platformy

2.3 Dátové zdroje a integrácia externých API

Kvalita a rozmanitosť vstupných údajov predstavujú základný predpoklad pre funkčnosť predikčných modelov. Súčasné finančné prostredie je charakteristické fenoménom 5Vs (*Volume, Variety, Velocity, Veracity, Value*), čo v praxi znemožňuje efektívne manuálne vyhodnocovanie všetkých relevantných udalostí [6]. Návrh FinPulse preto počíta s integráciou viacerých dátových prúdov, ktoré spoločne umožňujú modelovať trh nielen na základe historických cien, ale aj podľa sémantického obsahu aktuálneho diania. Bohatosť a rôznorodosť zdrojov, zahŕňajúca správy, makroekonomické ukazovatele aj emócie investorov, vytvára predpoklad pre presnejšiu identifikáciu budúcich cenových trendov [7].

2.3.1 Trhové a spravodajské dáta

Základným pilierom pre technickú analýzu a následné modelovanie cenových pohybov sú historické a reálne trhové údaje. FinPulse využíva rozhranie Yahoo Finance prostredníctvom knižnice `yfinance`, ktorá poskytuje prístup k štruktúrovaným dátam vo formáte OHLCV (Open, High, Low, Close, Volume). Tieto údaje sú nevyhnutné pre výpočet technických indikátorov a trénovanie modelov typu LSTM, ktoré sa dokážu učiť z časových závislostí v radoch [17].

Paralelne s číselnými údajmi systém využíva RSS kanály na zber finančného spravodajstva. Hoci priama korelácia medzi zverejnením správy a okamžitým cenovým pohybom nie je vždy jednoznačná, prehľady literatúry naznačujú, že textové informácie sú kľúčovým spúšťačom zvýšenej volatility a zmien v nálade trhových účastníkov [2]. Integrácia spravodajstva v navrhovanej architektúre slúži predovšetkým na zachytenie kontextu a identifikáciu významných udalostí, ktoré technické grafy samotné nedokážu zachytiť. Informácie z finančných zverejnení poskytujú dôležitý kontext o pozadí spoločností, čo je pre investičné rozhodovanie kritické [4].

2.3.2 Jazykové modely a sémantická analýza

Transformácia neštruktúrovaného textu správ do podoby využiteľnej pre algoritmy strojového učenia prebieha cez dva komplementárne prístupy. Prvým je integrácia rozhrania OpenAI API [26], kde model z rodiny GPT plní úlohu pokročilého sémantického extraktora a analytického poradcu. Jeho zapojenie v architektúre je navrhnuté na troch úrovniach:

- **Spracovanie a filtrovanie správ:** Každý prichádzajúci článok prechádza najprv posúdením relevantnosti. Model vyhodnocuje, či má text reálny zmysel pre finančnú analýzu, čím sa efektívne filtruje informačný šum [3]. Pri relevantných článkoch systém generuje štruktúrovaný výstup, ktorý obsahuje zhrnutie, sentimentové skóre, kategóriu udalosti a identifikáciu dotknutých titulov či sektorov. Návrh počíta aj s odhadom dĺžky dopadu správy a mierou dôvery modelu.

- **Analytický prehľad aktív:** Na úrovni detailu jednotlivých finančných inštrumentov systém využíva model na tvorbu textového súhrnu. Cieľom je poskytnúť investorovi rýchlu orientáciu v tom, čo pri danom aktíve sledovať a prečo je to dôležité. Výstup tak vytvára stabilný kvalitatívny rámec, ktorý prirodzene dopĺňa kvantitatívnu analýzu.
- **Inteligentný poradca portfólia:** Táto vrstva analyzuje agregované dáta z investičného účtu používateľa. Na základe celkového súhrnu pozícií, sektorového rozloženia a geografickej expozície model generuje prehľad silných a slabých stránok portfólia. Výsledkom sú aj konkrétne návrhy na diverzifikáciu a identifikáciu rizík. Výstup je navrhnutý tak, aby neobsahoval výpis jednotlivých aktív ani priame pokyny k nákupu či predaju, čím sa zachováva charakter poradenského nástroja.

Konkrétne znenie promptov pre všetky tri prípady použitia je v prílohe B.

Druhým prístupom je **sentimentová klasifikácia** textov správ predtrénovaným modelom FinBERT [14], ku ktorému FinPulse pristupuje cez knižnicu transformers. Model sa v rámci platformy nasadzuje len na inferenciu, bez dotrénovávania na vlastných dátach. Pre každý vstupný text článku priradí sentimentovú triedu a numerické skóre, ktoré sa normalizujú do rovnakej dátovej reprezentácie ako výstupy z vetvy GPT a z vlastného natrénovaného klasifikátora, aby ich bolo možné ukladať, porovnávať a zobrazovať v aplikácii jednotne. FinBERT nesie úlohu špecializovaného klasifikátora sentimentu, zatiaľ čo generatívna vetva GPT v návrhu rieši predovšetkým extrakciu meta-dát a štruktúrovanú analýzu. Podľa konfigurácie môže ísť o produkčný zdroj odhadu sentimentu alebo o záložnú vetvu pri nedostupnosti artefaktu vlastného modelu. Lokálna inferencia zároveň znižuje závislosť na externom API a latenciu pri dávkovom spracovaní správ.

2.3.3 Portfóliové a fundamentálne dáta

Aby platforma FinPulse poskytovala personalizovanú pridanú hodnotu, integruje rozhranie Trading212 API. Toto prepojenie umožňuje bezpečnú synchronizáciu investičného portfólia používateľa, vrátane

údajov o otvorených pozíciách a aktuálnom zisku či strate. Systém tak dokáže prioritne analyzovať informácie a generovať predikcie pre aktíva, ktoré majú priamy dopad na majetok konkrétneho investora.

Makroekonomický a korporátny kontext dodáva služba EarningsAPI [27]. Tento zdroj poskytuje kalendár hospodárskych výsledkov, informácie o dividendách a dôležité ekonomické udalosti. Tieto dáta sú podstatné, pretože moment zverejnenia finančných výsledkov býva obdobím, keď sa najvýraznejšie menia očakávania trhu a dochádza k revízii investičných stratégií [18]. Využitie týchto fundamentálnych vstupov v kombinácii s ostatnými zdrojmi vytvára ucelenú dátovú vrstvu, ktorá je nevyhnutným predpokladom pre spoľahlivé fungovanie celého systému.

2.4 Návrh ML pipeline

Proces strojového učenia je navrhnutý ako sekvenčný a reprodukovateľný tok dát, ktorý pokrýva celý životný cyklus modelu od zberu surových vstupov až po produkčnú inferenciu. Všeobecný rámec ML pipeline pre finančné aplikácie pozostáva zo zberu dát, predspracovania, výberu príznakov, rozdelenia dát, budovania modelu a jeho evaluácie [18]. Predkladaný návrh tento rámec rozširuje o hybridnú fúziu sentimentových a cenových príznakov a o cloudovú tréningovú infraštruktúru.

Pipeline je logicky rozdelená do dvoch vetiev. Prvá sa zameriava na analýzu sentimentu finančných správ, druhá na predikciu smeru pohybu ceny. V rámci tréningovej fázy boli experimentálne otestované varianty, v ktorých výstupy sentimentovej vetvy vstupujú ako príznaky do cenovej vetvy — výsledky týchto experimentov sú uvedené v kapitole 4. Pre produkčnú inferenciu bol zvolený model natrénovaný výlučne na cenových príznakoch, pretože historické správy nepokrývajú všetky sledované tituly a neumožňujú spoľahlivú fúziu s cenami naprieč celým rozsahom aktív.

2.4.1 Sentimentová vetva

Sentimentová vetva (*sentiment pipeline*) integruje tri klasifikačné prístupy hierarchicky usporiadané od jednoduchšieho ku komplexnej-

šiemu: logistickú regresiu ako štatistický baseline, neurónovú sieť LSTM a doménovo špecifický model FinBERT. Modely GPT z rodiny generatívnych transformerov plnia v rámci pipeline odlišnú úlohu — slúžia na extrakciu štruktúrovaných metadát, kategorizáciu typu udalosti a identifikáciu dotknutých aktív, nie na priamu sentimentovú klasifikáciu.

2.4.2 Predikčná vetva a evaluácia

Predikčná vetva smeru ceny (*price-move pipeline*) predstavuje finálny integračný bod systému. Na rozdiel od modelov pracujúcich výhradne s historickými cenami navrhovaná architektúra stavia na fúzii viacerých dátových prúdov. Denné sentimentové príznaky sa spájajú s cenovými príznakmi podľa dátumu. V prípadoch, kde pre daný deň nie sú dostupné žiadne správy, sa chýbajúce hodnoty dopĺňajú neutrálnou hodnotou, čím sa zabezpečuje konzistentnosť vstupnej schémy. Takáto kombinácia kvantitatívnych trhových signálov s kvalitatívnym informačným tokom zo správ umožňuje modelu reagovať na udalosti, ktoré ešte nemusia byť plne premietnuté v cenových indikátoroch [4].

Predikčný modul cenovej vetvy využíva tri architektúry: logistickú regresiu, LightGBM a LSTM. Dáta sa delia prísne časovo na tréningovú, validačnú a testovaciu množinu bez akéhokoľvek miešania časovej osi, čo eliminuje problém úniku informácií z budúcnosti [4]. Návrh zároveň počíta aj s variantom pipeline bez sentimentu (*price-only*), čo umožňuje experimentálne overenie prínosu sentimentovej vetvy.

Výsledkom celého procesu je binárna klasifikácia smeru pohybu ceny (**UP/DOWN**) pre nasledujúci obchodný deň. Spolu s modelom sa ukladajú metadáta obsahujúce zoznam použitých príznakov, parametre a metriky, čo zaručuje, že produkčná inferencia používa identickú vstupnú schému ako tréning.

2.4.3 Cloudová infraštruktúra a nasadenie

Vzhľadom na výpočtovú náročnosť tréningovania hlbokých neurónových sietí, ako sú modely LSTM [17], je tréningová fáza delegovaná do cloudového prostredia Microsoft Azure Machine Learning [28]. Tento prístup umožňuje FinPulse využívať výkon vzdialených výpočtových clusterov bez dopadu na odozvu produkčnej aplikácie. Po úspešnom

natrénovaní a vyhodnotení metrík sú modelové artefakty prenesené do lokálneho úložiska backendu, kde prebieha samotná inferencia — čím sa zachováva nízka latencia pri každodennom používaní systému.

2.5 Návrh používateľského rozhrania

Používateľské rozhranie FinPulse má slúžiť ako jeden pracovný priestor, kde investor na jednom mieste sleduje trhové signály, sentiment aj spravodajstvo a prirodzene prechádza do detailnejších analytických modulov. Klientska časť je navrhnutá ako *Single Page Application* s technologickým *stackom* React, TypeScript, Vite, Tailwind CSS a React Router. Rozloženie rozhrania s horným panelom, bočnou navigáciou a centrálnou oblasťou obsahu zodpovedá potrebe mať všetky moduly na jednom mieste.

Prehľadový dashboard zobrazuje agregovaný sentiment nad spracovanými správami, cenové pohyby sledovaných titulov a rebríčky denných výhercov a porazených. Hlavný spravodajský kanál umožňuje filtrovanie podľa kategórie správy alebo aktíva a prepínanie medzi celkovým trhom a osobným *watchlistom*.

Sledovanie aktív a detail Používateľ spravuje *watchlist* sledovaných titulov. V detaile každého aktíva je dostupný AI súhrn silných stránok, slabín a rizík a prehľad hospodárskych výsledkov spoločnosti.

Portfólio a AI poradca Po prepojení cez Trading212 API systém zobrazuje aktuálne pozície, sektorové a geografické rozloženie a metriky zisku a straty. AI modul na základe zloženia portfólia identifikuje koncentračné riziká a navrhuje smery diverzifikácie.

Analytické moduly Samostatné obrazovky zobrazujú index cieľových cien inštitucionálnych analytikov, ekonomický kalendár a prehľad trhových kapitalizácií.

Snímky hlavných obrazoviek sú v prílohe A.

3 Implementácia

Realizačná fáza priamo nadväzuje na návrhové rozhodnutia opísané v predchádzajúcej kapitole a transformuje ich do funkčného softvérového celku. Cieľom implementácie bolo vybudovať FinPulse ako systém schopný asynchrónne spracovávať heterogénne dátové toky, vykonávať inferenciu natrénovaných modelov a prezentovať výsledky analytiky prostredníctvom interaktívneho rozhrania. Nasledujúce podkapitoly detailne opisujú realizáciu jednotlivých komponentov — od prípravy vstupných dát cez tréning modelov strojového učenia až po implementáciu serverovej a klientskej vrstvy FinPulse.

3.1 Dátová príprava a predspracovanie

Vo FinPulse prebieha príprava dát v samostatnom ML module, ktorý zabezpečuje tok dát od príjmu surových vstupov po generovanie finálnych príznakov pre obe vetvy pipeline. Celý proces sa skladá z troch na seba nadväzujúcich fáz: predspracovania textových správ, predspracovania cenových dát a ich záverečnej fúzie.

Vstupom sú surové správy vo formáte CSV pochádzajúce z verejne dostupného datasetu historických finančných správ o akcii Apple [29], pričom sa očakávajú stĺpce pre dátum, nadpis a obsah článku. Prvým krokom je normalizácia dátumov na dennú úroveň, čo zabezpečuje konzistentnú granularitu pri následnej agregácii. Ak v zdrojových dátach chýbajú sentimentové stĺpce, doplnia sa nulové hodnoty, čím sa zachováva stabilita vstupnej schémy.

Z nadpisu a obsahu správy sa pomocou pravidlových mechanizmov odvodzuje typ udalosti (`event_type`). Kategorizácia rozlišuje napríklad `earnings`, `product_launch`, `regulatory`, `macro` a zástupnú triedu `other`.

Každá správa v CSV obsahuje stĺpec `symbols`, ktorý uvádza burzové symboly (`tickery`) súvisiace s daným článkom. Pri výpočte sentimentových príznakov pre konkrétny ticker sa najprv vyberú iba tie riadky, v ktorých `symbols` obsahuje hľadaný ticker ako podreťazec. Tým sa každý článok jednoznačne priradí k sledovanému titulu ešte pred agregáciou. Vybrané záznamy sa následne zoskupia podľa

kalendárneho dňa a pre každý deň sa vypočíta priemerná polarita a sentimentové skóre, počty spracovaných článkov a distribúcia identifikovaných typov udalostí. Na základe týchto primárnych príznakov sa dopočítavajú rozšírené sentimentové atribúty — konkrétne `attention_change` (zmena pozornosti médií) a `abnormal_sentiment` ako odchýlka aktuálneho sentimentu od jeho kľzavého priemeru. Výsledný dataset sentimentových príznakov sa ukladá ako CSV súbor pre ďalšie spracovanie.

Historické cenové dáta vo formáte OHLCV (Open, High, Low, Close, Volume) sa získavajú pre zvolené akciové tituly prostredníctvom knižnice `yfinance`, ktorá sprístupňuje historické trhové dáta z Yahoo Finance. Z týchto surových hodnôt sa generuje sada technických a štatistických príznakov:

- **Výnosové príznaky a momentum:** denné výnosy, posuny toho istého radu o jeden až viac obchodných dní do minulosti (*lagové* príznaky) a krátkodobé momentum zachytávajú dynamiku a rýchlosť cenových zmien v čase.
- **Trendové príznaky (MA):** kľzavé priemery v rôznych časových oknách a ich vzájomné pomery umožňujú identifikovať prevládajúci smer pohybu aktíva.
- **Volatilita a štatistiky v posuvnom okne:** miery variability a kľzavé priemery či smerodajné odchýlky výnosov v oknách niekoľkých dní zachytávajú úroveň trhovej neistoty.
- **RSI a MACD:** RSI meria rýchlosť a intenzitu cenových pohybov, zatiaľ čo MACD sleduje vzájomný vzťah dvoch exponenciálnych kľzavých priemerov a ich signálovú líniu pre identifikáciu trendových zlomov.
- **Objemové príznaky:** normalizované zmeny a pomery objemov obchodov slúžia na potvrdenie sily identifikovaného pohybu.

Okrem príznakov sa v tejto fáze generuje aj predikčný cieľ (*target*). Konkrétne sa vypočíta `next_return_1d` (výnos nasledujúceho dňa), binárny label UP/DOWN a indikátor `in_band` pre identifikáciu plochého trhového pohybu. Riadky s neplatnými hodnotami sú z dát vyčistené,

pričom pre potreby inferencie je k dispozícii režim `keep_targetless_tail`, ktorý zachová posledné riadky aj bez dostupného cieľa. Výsledok sa ukladá ako štruktúrovaný CSV súbor cenových príznakov.

Záverečným krokom prípravy dát je fúzia oboch spracovaných datasetov. Pri aktivovanom režime `use-sentiment` sa k tabuľke cenových príznakov pre zvolený ticker pripája tabuľka denných sentimentových príznakov, ktorá bola v predchádzajúcom kroku už vyfiltrovaná a agregovaná pre ten istý ticker. Spojenie prebieha podľa dátumu obchodného dňa, pričom primárnou tabuľkou sú ceny, takže v množine ostávajú všetky obchodné dni z cenového radu aj pre dni bez správ. Pre takéto dni sa chýbajúce hodnoty sentimentových príznakov dopĺňajú nulou, čím sa zabezpečuje konzistentnosť vstupnej schémy a zamedzuje sa narušeniu tréningového procesu.

Pri tréňovaní modelu sa časová postupnosť pozorovaní delí na tréningovú, validačnú a testovaciu množinu bez miešania poradia na časovej osi.

3.2 Architektúra a tréňovanie modelov analýzy sentimentu

Analýza sentimentu finančných správ je implementovaná v samostatnom ML module, ktorý pokrýva načítanie dát, tréning, uloženie modelov a inferenciu. Tréningová množina vychádza z datasetu Financial PhraseBank [30] — zbierky 4 846 viet z finančných správ v anglickom jazyku, ktoré kategorizovalo 5 až 8 anotátorov podľa sentimentu — kde každá vzorka má formát `sentence@label` s triedami `positive`, `neutral` a `negative`. Pri načítaní prebieha validácia vstupov — overenie existencie súboru, kontrola formátu riadkov a vynechanie neplatných záznamov. Dáta sú následne rozdelené na tréningovú, validačnú a testovaciu množinu so zachovaním pôvodného zastúpenia tried v každej časti datasetu. Experimentálne prostredie je postavené na knižniciach `scikit-learn`, `torch`, `transformers`, `joblib`, `numpy` a `pandas`.

Logistická regresia je implementovaná ako klasifikačná pipeline kombinujúca TF-IDF vektorizáciu s lineárnym klasifikátorom. Vstupné texty sú transformované na numerickú reprezentáciu pomocou unigra-

mov a bigramov. Keďže dataset obsahuje nerovnomerné zastúpenie tried, klasifikátor trénuje s vyvažovaním tried. Natrénovaný model sa ukladá spolu s metrikami (accuracy, macro-F1, konfúzna matica) a tréningovými parametrami. Konkrétne nastavenia vektorizátora: `max_features=20 000` a `min_df=2`. Klasifikátor používa parametre `max_iter=2000` a `class_weight="balanced"`. Model sa ukladá vo formáte `.joblib`, metriky a parametre behu v pridruženom `.meta.json`.

LSTM pristupuje k textu ako k sekvencii tokenov. Implementácia využíva vlastnú tokenizáciu a slovník budovaný z tréningových textov, pričom dlhšie texty sú orezané na pevnú maximálnu dĺžku. Architektúra pozostáva z embeddingovej vrstvy, rekurentnej vrstvy LSTM, regularizácie typu *dropout* a záverečnej lineárnej klasifikačnej vrstvy. Tréning prebieha optimalizátorom Adam s cross-entropy stratou, veľkosť dávky a počet epoch sú konfigurovateľné. Maximálna dĺžka sekvencie je `max_len=128` tokenov, veľkosť dávky `batch_size=16` a počet epoch je 50. Po tréningu sa model ukladá ako `.pt` checkpoint so stavom váh a mapovaniami `word2idx` a `label2id`. Metriky a parametre behu sú v sprievodnom `.meta.json`.

FinBERT je v systéme nasadený ako inferenčný model bez lokálneho dotrénovávania, pričom sa načítava predtrénovaný checkpoint ProsusAI/finbert cez knižnicu transformers. V implementácii tvorí samostatnú vetvu sentimentovej inferencie, ktorá pre každý vstupný text správy vracia normalizovaný výstup vo forme sentimentového štítku a numerického skóre. Táto vetva je integrovaná do spoločného API rozhrania spolu s GPT a vlastným modelom, takže výsledky všetkých modelov sú spracované v rovnakom formáte pre následné porovnanie a zobrazenie v klientskej vrstve. Ak je nakonfigurovaný vlastný trénovaný model (LR alebo LSTM), ale jeho súbor v adresári `ml/artifacts/` chýba, inferencia automaticky prepne na FinBERT.

GPT je volaný cez OpenAI API v dvoch fázach (presné znenie promptov je v prílohe B). Najprv sa krátkym promptom s titulkom a úryvkom tela overí, či je článok relevantný pre finančnú analýzu — tento filter znižuje počet plných API volaní. Pre relevantné články nasleduje hlavná analýza, ktorej vstup tvorí titulok a telo orezané na

približne 2000 znakov. Výstup je štruktúrovaný JSON so siedmimi poľami: titulok, sumarizácia, sentimentová kategória a skóre, zoznam dotknutých aktív s odhadom smeru dopadu, časový horizont dopadu a miera istoty. Rovnaké API rozhranie sa používa aj pri generovaní investičných súhrnov v detaile aktíva a pri analýze portfólia.

3.3 Konštrukcia price-move modelov a fúzia dát

Predikcia smeru pohybu ceny je implementovaná v samostatnej vetve, ktorá zdieľa rovnakú infraštruktúru pre správu artefaktov a metadát so sentimentovou vetvou. Modul realizuje binárnu klasifikáciu smeru pohybu ceny (UP/DOWN) pre nasledujúci obchodný deň z denných cenových dát vo formáte OHLCV. Výhodou tohto prístupu je, že pipeline môže fungovať buď samostatne s čisto cenovými vstupmi, alebo v kombinovanom režime so sentimentovými príznakmi.

Vstupné dáta sú získavané pre zvolený ticker cez knižnicu `yfinance` alebo načítavané z lokálneho úložiska. Predikčný cieľ je odvodený z hodnoty `next_return_1d` — ak je výnos nasledujúceho dňa väčší alebo rovný nule, dňu sa priradí label UP, v opačnom prípade DOWN. Dni s malým pohybom pod definovaným prahom sú označené ako `in_band` a pred časovým rozdelením na trénovaciu, validačnú a testovaciu množinu sa z datasetu vynechajú. Tento mechanizmus zabezpečuje, že model netrénuje na trhovom nevýznamných pohyboch, ktoré by mohli zavádzať jeho rozhodovaciu logiku. Pre inferenciu posledného dostupného dňa sa využíva špeciálny režim `keep_targetless_tail`, ktorý zachová riadok aj bez dostupného cieľa.

Tréning modelov využíva knižnice `scikit-learn`, `torch`, `joblib` a `lightgbm` a podporuje tri architektúry:

Logistická regresia je implementovaná ako pipeline skladajúca sa z normalizácie príznakov a lineárneho klasifikátora. Prípadné chýbajúce hodnoty príznakov sú pred normalizáciou nahradené mediánom z tréningových dát. Keďže zastúpenie tried UP a DOWN v historických dátach je prirodzene nerovnomerné, klasifikátor trénuje s vyvažovaním tried (`class_weight="balanced"`).

LightGBM je tréňovaný priamo nad vektorom cenových príznakov ako gradientný boosting klasifikátor. Vyvažovanie tried kompenzuje nerovnomerné zastúpenie UP a DOWN v historických dátach. Na predchádzanie pretrénovania sa počas tréňovania monitoruje výkon na validačnej množine a tréning sa predčasne zastaví, ak sa metrika nezlepší. Konkrétne hodnoty: `n_estimators=300`, `learning_rate=0,05`, `num_leaves=31`. Parametre vzorkovania: `subsample=0,8`, `colsample_bytree=0,8`, `class_weight="balanced"`. Skoré zastavenie sleduje metriku `binary_logloss` a nastáva po 50 iteráciách bez zlepšenia.

LSTM pristupuje k predikcii ako k problému na sekvenčných dátach — namiesto jedného izolovaného dňa dostáva na vstupe časové okno predchádzajúcich obchodných dní. Architektúra sa skladá z jednej rekurentnej vrstvy a výstupnej lineárnej vrstvy pre binárnu klasifikáciu. Tréning prebieha optimalizátorom Adam. Dĺžka vstupného okna je `lookback=20` obchodných dní, skrytá vrstva má `hidden_size=64` jednotiek, rýchlosť učenia `lr=0,0001` a tréning prebieha počas 30 epoch. Po tréningu je model ukladán ako `.pt` checkpoint.

Dáta sú rozdeľované striktne temporálne na tréňovaciu, validačnú a testovaciu množinu. Pri tréningu na viacerých tickeroch súčasne prebieha split samostatne pre každý ticker a výsledné množiny sa následne spájajú (pooling), čím sa zachováva časová integrita každého aktíva. Vyhodnotenie modelov prebieha cez *balanced accuracy*, *macro-F1* a *konfúznú maticu*. Spolu s modelom sa ukladajú metadáta, ktoré zaručujú, že produkčná inferencia vždy použije identickú vstupnú schému ako tréning. Metadáta sú uložené vo formáte `.meta.json` a obsahujú zoznam použitých príznakov, tréningové parametre a metriky.

Pri aktivovanom režime `-use-sentiment` sa k cenovým príznakom pre daný ticker pripájajú sentimentové príznaky rovnakým spôsobom ako pri fúzii popísanej vyššie: filtrovanie podľa stĺpca `symbols`, denná agregácia a spojenie podľa dátumu s cenovou tabuľkou ako primárnou. Kombinácia viacerých titulov a sentimentového režimu súčasne nie je podporovaná.

Inferenčný skript prijíma CSV súbor s vopred vypočítanými príznakmi a vracia štruktúrovaný výstup vo formáte JSON obsahujúci smer pohybu (`direction`: UP alebo DOWN), pravdepodobnosť rastu

(`probability_up`, ak model podporuje pravdepodobnostné odhady) a referenčný dátum predikcie (`as_of_date`).

V produkčnom prostredí je predikcia integrovaná do aplikačnej vrstvy prostredníctvom FastAPI endpointu. Aplikačná vrstva orches-truje celý inferenčný tok ako sekvenciu podprocesov: stiahnutie aktuálnych cenových dát, výpočet príznakov a volanie natrénovaného modelu. Pre skupinu sledovaných aktív je nasadený fixný artefakt LightGBM natrénovaný na viacerých tituloch naraz, pričom výsledky predikcií sú kešované v databáze podľa tickeru a dátumu. Tento prístup minimalizuje redundantné výpočty a zaručuje rýchlu odozvu klientskej aplikácie pri opakovaných dotazoch na rovnaké aktíva.

3.4 Serverová logika a konektivita externých služieb

Aplikačná vrstva na strane servera je implementovaná ako asynch-rónna FastAPI aplikácia. Jej úlohou nie je trénovanie modelov, ale orchestrácia dátových tokov, volanie inferenčných skriptov a obsluha REST API pre klientsku časť. Kód je organizovaný do dvoch hlavných vrstiev. Smerovacia vrstva v `app/api/` definuje endpointy rozčlenené do modulov podľa funkčných celkov (správy, trh, portfólio, hospo-dárske výsledky, aktíva a autentifikácia), zatiaľ čo aplikačná logika je zapuzdrená v servisnej vrstve `app/services/`. Konfigurácia aplikácie vrátane API kľúčov a prepínačov externých integrácií je spravovaná prostredníctvom `.env` súboru a centrálného konfiguračného objektu `settings`, čo umožňuje bezpečné nasadenie bez hardkódovaných cit-livých hodnôt v zdrojovom kóde. Všetka odchádzajúca HTTP komuni-kácia s externými rozhraniami prebieha asynchrónne prostredníctvom knižnice `httpx`.

Servisná vrstva je rozdelená do niekoľkých funkčne samostatných tried, z ktorých každá zodpovedá za jednu doménu aplikačnej logiky.

MarketDataService zabezpečuje získavanie trhových dát prostred-níctvom knižnice `yfinance`. Načítava živé ceny vybraných makroeko-nomických ukazovateľov (S&P 500, Bitcoin, zlato, index amerického dolára DXY) a zostavuje zoznamy najväčších denných nárastov a po-klesov prostredníctvom Yahoo Finance screenerov. Pre výpočet refe-renčnej hodnoty *fair value* volá metódu `get_analyst_price_targets()`,

ktorá vracia priemerný a mediánový cieľový kurz analytikov. Výsledky sú udržiavané v pamäťovej cache, pričom čas platnosti záznamu sa líši podľa rýchlosti zmeny dát, čím sa vyvažuje aktuálnosť voči záťaži externého rozhrania. Konkrétne hodnoty TTL: živé ceny 1 hodina, trendy 12 hodín, *fair value* 24 hodín.

NewsIngestionService spravuje príjem finančných správ z RSS kanálov Yahoo Finance, pričom záznamy článkov sa z kanálovej odpovede získavajú parsovaním prostredníctvom knižnice *feedparser*. Každý záznam je pred uložením skontrolovaný na duplicitu podľa URL adresy, čím sa predchádza opakovanému spracovaniu tých istých článkov. Nové záznamy sú uložené do databázy SQLite, odkiaľ ich preberá pipeline AI analýzy.

AI SummarizationService implementuje dvojfázové spracovanie finančných správ. V prvej fáze model GPT rozhodne, či je článok relevantný pre finančnú analýzu. Táto kontrola využíva kratší prompt s nižšou spotrebou tokenov, čím sa minimalizujú náklady na API volania. V druhej fáze je pre relevantné články vykonaná plná analýza, ktorej výstup je štruktúrovaný JSON so siedmimi poľami: titulok, sumarizácia, sentimentové skóre, kategória udalosti, dotknuté tituly vrátane smeru dopadu, časový horizont a úroveň istoty predikcie.

AssetSummaryService a PortfolioAnalysisService sú dve služby postavené na rovnakom OpenAI rozhraní, avšak s odlišným zameraním. **AssetSummaryService** generuje stručný investičný súhrn pre zadané aktívum štruktúrovaný do troch častí: aktuálne dianie, zdôvodnenie relevantnosti pre investora a body vyžadujúce sledovanie. Prompt je explicitne obmedzený na 150 slov a zakazuje modelu uvádzať hodnoty, ktoré nie sú súčasťou vstupného kontextu, čím sa predchádza halucinácii. Súhrn je podložený aktuálnymi webovými zdrojmi prostredníctvom OpenAI Responses API s nástrojom *web_search*. **PortfolioAnalysisService** zostavuje z dát portfólia štruktúrovaný kontext zahŕňajúci celkovú hodnotu, nerealizovaný zisk a stratu a percentuálne sektorové a geografické rozloženie. Na jeho základe model vracia JSON s piatimi poľami: celkové odporúčanie, silné stránky, slabiny, riziká a návrhy na diverzifikáciu. Model je inštruovaný, aby

výstup obsahoval výlučne poradenskú zložku bez opakovania vstupných údajov o pozíciách.

PortfolioService zabezpečuje synchronizáciu portfólia z Trading212 API. Z každej pozície sú extrahované kvantitatívne údaje: množstvo, priemerná nákupná cena, aktuálna trhovú hodnota, nerealizovaný zisk a strata a devízový dopad. Sektorové a geografické zaradenie jednotlivých pozícií nie je súčasťou Trading212 odpovede a je dohľadane dodatočne prostredníctvom Yahoo Finance na základe ISIN kódu každého aktíva. V prípade potreby menovej konverzie je výmenný kurz taktiež načítaný z `yfinance`.

EarningsService agreguje štyri typy dát z EarningsAPI: kalendár hospodárskych výsledkov s rozdelením podľa času zverejnenia (pred otvorením a po zatvorení trhu), dividendový kalendár, makroekonomický kalendár udalostí a plné textové súhrny výsledkov pre jednotlivé spoločnosti. Všetky odpovede sú ukladané do pamäťovej cache s TTL 24 hodín.

SentimentService zabezpečuje ML sentimentovú predikciu pre spracované správy. Služba podporuje tri zdroje sentimentu konfigurovateľné prostredníctvom `settings.SENTIMENT_SOURCE`: GPT (predvoľené), FinBERT a vlastný trénovaný ML model. Pri zdrojoch FinBERT a vlastnom modeli táto služba pre každý vstupný text vytvorí dočasný súbor, zavolá skript `predict_sentiment` ako podproces a výsledok prečíta zo štandardného výstupu. Dočasný súbor je po spracovaní okamžite zmazaný. Pri zdroji GPT sa sentiment a skóre preberajú priamo z výsledku AI analýzy článku bez volania externého skriptu. Tým je dosiahnutá flexibilita pri výbere kompromisu medzi rýchlosťou, nákladmi a presnosťou. Podproces je ohraničený timeoutom 60 sekúnd.

PriceMoveService implementuje predikciu smeru pohybu cien pre skupinu siedmich technologických titulov (MAG7). Pri každej požiadavke táto služba najprv overí, či v databáze už existujú uložené predikcie pre všetky požadované tickery a posledný uzavretý obchodný deň. Ak áno, vráti ich bez spustenia ďalších výpočtov. Tým sa znižuje

počet výpočtovo náročných behov inferenčnej pipeline pri súčasnom sledovaní viacerých titulov. V prípade cache miss pipeline sekvenčne stiahne historické cenové dáta, vypočíta technické príznaky a zavolá natrénovaný model, ktorý vráti predikciu vo formáte JSON vrátane smeru pohybu a pravdepodobnosti nárastu. Výsledky sú uložené do databázy SQLite prostredníctvom SQLAlchemy ORM. Konkrétne ide o sekvenčné volanie skriptov `fetch_price`, `build_price_features` a `predict_price_move`, pričom každý podproces je ohraničený `time-out`om 120 sekúnd.

Výpočtovo náročný tréning, najmä LSTM (sentiment alebo smer pohybu ceny) a LightGBM natrénovaný na viacerých tituloch v rámci vetvy pohybu ceny, je možné delegovať do prostredia Microsoft Azure Machine Learning skriptom `submit_azure_job.py`. Skript odošle príkazovú úlohu na zvolený výpočtový cluster a v jej konfigurácii nastaví parametre vrátane typu úlohy (`-job`), tickeru alebo zoznamu tickerov pri LightGBM, voliteľného dátumového ohraničenia (`-start-date`, `-end-date`) a príznaku sentimentového režimu (`-use-sentiment`). Po stiahnutí výstupov z Azure ML sa modely uložia do príslušných podadresárov v `ml/artifacts/`, odkiaľ `PriceMoveService` pri produkcii načítava napríklad model pohybu cien natrénovaný na viacerých tituloch.

Kľúčovým implementačným rozhodnutím spoločným pre inferenciu sentimentu a predikciu pohybu cien je integrácia prostredníctvom volania podprocesov príkazmi `python -m ml.scripts...` z koreňa repozitára. Tento prístup udržiava balík `ml` oddelený od procesu aplikačného servera a umožňuje nasadenie bez potreby inštalovať celý ML ekosystém do prostredia aplikačného servera.

Bezpečnosť API je riešená na dvoch úrovniach: ochrana hesiel a relačná autentifikácia. Heslá sa neukladajú otvorene — pri registrácii sa z nich vytvára haš algoritmom `bcrypt` prostredníctvom knižnice `passlib` a pri prihlásení sa overujú oproti uloženému hašu. Relácie používajú JWT tokeny, ktoré sa po úspešnej registrácii alebo prihlásení ukladajú do HTTP cookie s názvom `access_token`. Chránené endpointy pri každej požiadavke overia JWT z cookie a pri platnom tokene vyhládajú v databáze zodpovedajúceho používateľa. FastAPI podporuje asynchrónne handlersy a neblokujúce sieťové volania, čím sa znižuje riziko, že dlhšie operácie budú zdržiavať obsluhu ďalších

súbežných požiadaviek, pričom podprocesy ML sú v službách ohraničené timeoutmi.

3.5 Klientska aplikácia a vizualizácia dát

Klientska vrstva prezentuje dáta z REST API serverovej aplikácie a zabezpečuje interaktívne ovládanie funkcií vyžadujúcich autentifikáciu. Inferenciu modelov ani priamy prístup k externým dátovým zdrojom nevykonáva. Výpočty, validácie a napojenie na služby tretích strán sú výhradne na strane servera.

Zdrojový kód klienta v `web/src/` rozdeľuje zodpovednosti medzi priečinky `components/`, `pages/`, `hooks/`, `lib/` a `context/`.

Koreňový komponent obaľuje smerovanie do `ErrorBoundary`, ktorý zachytáva neočakávané chyby vykresľovania a zabraňuje pádu celej stránky. Počiatočná obrazovka zobrazuje krátku animáciu načítania s identitou aplikácie, čím sa zjednotí prvý dojem pred zobrazením hlavného rozloženia.

Rozloženie a smerovanie Hlavné rozhranie je postavené na komponente `MainLayout` kombinujúcom `Sidebar`, `Navbar` a výstupný priestor pre aktívnu stránku. Trasy sú definované deklaratívne. `Protected Route` vynucuje prihlásenie pred vstupom do chránených sekcií a ne-prihláseného používateľa presmeruje na prihlásenie. Neexistujúca URL adresa končí na stránke chyby 404.

Komunikácia so serverom Trieda `ApiClient` centralizuje odosielanie požiadaviek prostredníctvom štandardného rozhrania `fetch`. Konštruje URL vrátane query parametrov, očakáva odpovede vo formáte JSON a pri chybových HTTP stavoch mapuje telo odpovede na jednotný formát chyby. Pri každej požiadavke je nastavené `credentials: include`, čo umožňuje prenášať HTTP cookie s JWT tokenom vydaným serverom bez manuálnej správy hlavičky `Authorization`.

Autentifikácia a globálny stav Kontext `AuthContext` uchováva informáciu o prihlásenom používateľovi, načítava ju endpointom `/api/auth/me` pri štarte aplikácie a poskytuje metódy na registráciu, prihlásenie, odhlásenie a úpravu profilu. Modálne okná na prihlásenie

a registráciu sú oddelené do vlastného kontextu, aby boli dostupné z viacerých miest rozhrania bez duplikácie kódu. Systém toast notifikácií informuje používateľa o úspešných alebo neúspešných operáciách.

Načítavanie dát Pre jednotlivé domény (trhové dáta, správy, portfólio, kalendáre, predikcie) existujú vlastné React hooky v priečinku `hooks/`. Tým sa oddeľuje volanie API od prezentačných komponentov, zjednocuje sa spracovanie stavov načítavania a chýb a znižuje sa opakovanie kódu medzi stránkami.

Vizualizácia dát Okrem tabuliek, kariet a zoznamov využíva aplikácia knižnicu `Recharts`. Pre vizualizáciu rozloženia portfólia sú implementované interaktívne koláčové diagramy (`PieChart`). Komponent `ResponsiveContainer` zabezpečuje prispôsobenie rozmerov grafov šírke kontajnera pri rôznych veľkostiach okna prehliadača. Príklady vizualizácií sú v prílohe A.

Horný panel a doplnkové akcie V `Navbar` je integrované vyhľadávanie aktív s automatickým dopĺňaním, ktoré po výbere presmeruje na stránku detailu. Tlačidlá na manuálne spustenie načítania správ a AI analýzy volajú príslušné REST endpointy servera a umožňujú obnovenie obsahu na požiadanie priamo z rozhrania. Prepínač svetlého a tmavého motívu mení atribút `data-theme` na koreňovom elemente dokumentu a ukladá preferenciu do `localStorage`, čím sa zachová voľba medzi reláciami.

4 Experimenty a vyhodnotenie

Táto kapitola predstavuje experimentálne overenie navrhnutých modelov strojového učenia na reálnych dátach. Cieľom experimentov je vyhodnotiť a porovnať sentimentové modely a modely pre predikciu cenových pohybov a overiť, či kombinácia sentimentových príznakov s technickými indikátormi prináša merateľné zlepšenie oproti predikcii postavenej výhradne na historických cenových dátach.

Výsledky sú hodnotené pomocou štandardných metrík strojového učenia, pričom dôraz je kladený nielen na štatistickú presnosť, ale aj na interpretáciu výsledkov v kontexte finančnej analytiky.

4.1 Datasets a experimentálny protokol

Táto podkapitola opisuje použité datasets, spôsob rozdelenia dát a voľbu hodnotiacich metrík. Uvedené nastavenia platia pre všetky experimenty v podkapitolách 4.2 a 4.3.

Sentiment Použitý dataset je Financial PhraseBank, konkrétne variant Sentences_50Agree s 4 846 vetami anglického finančného spravodajstva. Dataset je výrazne nevyvážený: trieda `neutral` tvorí 2 566 príkladov (52,9 %), `positive` 1 549 (32,0 %) a `negative` 731 (15,1 %). Rozdelenie na tréning, validáciu a test je stratifikované v pomere 70/15/15, aby každá časť zachovala pôvodné zastúpenie tried. Parameter `seed = 1` zabezpečuje reprodukovateľnosť rozdelenia. Model FinBERT nie je trénovaný, slúži výhradne ako referenčný bod pri inferencii.

Predikcia pohybu ceny Vstupom sú denné cenové príznaky (OHLCV, technické indikátory) pre každý ticker. Rozdelenie je časové v pomere 70/15/15 bez náhodného miešania, čím sa zachováva chronologický poriadok. Pri tréningu na viacerých tituloch (MAG7) sa časový rez počíta pre každý symbol samostatne a množiny sa spoja. Kombinácia cenových príznakov so sentimentom je z dátových dôvodov podporovaná len pre jeden ticker naraz.

Metriky Pre sentiment sa používa presnosť a makro-F1. Makro-F1 je uprednostnené pred celkovou presnosťou, pretože priraduje všetkým triedam rovnakú váhu nezávisle od ich početnosti. Pre predikciu pohybu ceny sa používa vyvážená presnosť a makro-F1 — vyvážená presnosť je zvolená preto, že triedy UP a DOWN nemusia byť v dátach rovnomerne zastúpené a bežná presnosť by favorizovala majoritnú triedu.

4.2 Vyhodnotenie sentimentových modelov

Táto podkapitola prezentuje výsledky klasifikácie sentimentu na datase Financial PhraseBank s rozdelením a metrikami opísanými v podkapitole 4.1. Kvantitatívne sú porovnávané dva vlastné modely zo skriptu `train_sentiment`: logistická regresia nad TF-IDF vektormi (ďalej LR) a rekurentná sieť LSTM.

4.2.1 Kvantitatívne výsledky

LR aj LSTM boli natréňované s identickým stratifikovaným rozdelením a hodnotené na rovnakej validačnej a testovacej množine. Výsledky sú zhrnuté v tabuľke 4.1.

Tabuľka 4.1: Porovnanie sentimentových modelov na validačnej a testovacej množine.

Model	Val presnosť	Val makro-F1	Test presnosť	Test makro-F1
TF-IDF + LR	0,747	0,714	0,783	0,763
LSTM	0,722	0,685	0,729	0,701

Na čistom datase s 727 testovacími príkladmi predbieha LR model LSTM o 5,4 percentuálneho bodu v presnosti a o 6,2 bodu v makro-F1. Oba modely dosahujú hodnoty výrazne nad úrovňou náhodného klasifikátora (makro-F1 $\approx 0,33$ pri troch triedach). Testovacia množina nie je vyvážená — trieda `neutral` tvorí väčšinu príkladov (385 z 727) — preto makro-F1, ktoré priraduje všetkým triedam rovnakú váhu, lepšie odráža výkon na menej zastúpených triedach (`negative`, `positive`) než celková presnosť.

4.2.2 Analýza chýb

Matice zámen oboch modelov na testovacej množine (riadky = skutočná trieda, stĺpce = predikovaná):

$$\text{LR: } \begin{pmatrix} 77 & 19 & 13 \\ 8 & 331 & 46 \\ 13 & 59 & 161 \end{pmatrix} \quad \text{LSTM: } \begin{pmatrix} 74 & 23 & 12 \\ 24 & 312 & 49 \\ 19 & 70 & 144 \end{pmatrix}$$

Poradie riadkov aj stĺpcov: negative, neutral, positive

Oba modely najčastejšie zamieňajú triedy positive a neutral. LR zaznamenáva 59 zámen positive za neutral a 46 zámen neutral za positive. LSTM zaznamenáva 70 a 49 takýchto zámen. Hranica medzi positive a neutral je prirodzene nejasná, pretože pozitívne správy často používajú lexiku, ktorá vyzerá neutrálne.

Trieda negative je klasifikovaná presnejšie než ostatné, čo pravdepodobne súvisí so špecifickejšou lexikou poklesu a strát v tréningových dátach. Podrobné metriky po triedach uvádza tabuľka 4.2.

Tabuľka 4.2: Metriky po triedách na testovacej množine.

Model	Trieda	P	R	F1	<i>n</i>
TF-IDF + LR	negative	0,786	0,706	0,744	109
	neutral	0,809	0,860	0,834	385
	positive	0,732	0,691	0,711	233
LSTM	negative	0,633	0,679	0,655	109
	neutral	0,770	0,810	0,790	385
	positive	0,702	0,618	0,658	233

P = presnosť (precision), R = úplnosť (recall), *n* = počet príkladov.

LR dosahuje lepšie výsledky vo všetkých triedach. Najväčší rozdiel je pri triede negative (F1 LR = 0,744 vs. LSTM = 0,655) a positive (0,711 vs. 0,658). Prekvapivé poradie modelov možno vysvetliť veľkosťou tréningovej sady: s 3 392 tréningovými príkladmi nemusí mať LSTM dostatok dát na efektívne naučenie sekvenčných závislostí a TF-IDF reprezentácia zachytí kľúčovú doménovú lexiku aj pri menšom objeme dát.

4.2.3 Vzťah k modelu FinBERT

Model FinBERT bol použitý ako referenčný bod. Predtrénovaný checkpoint ProsusAI/finbert bol pri publikácii modelu fine-tunovaný okrem iného aj na Financial PhraseBank. Pri porovnaní jeho predikovaných štítkov s ľudskými anotáciami na tom istom korpuse vychádza 100 % zhoda a matica zámen je diagonálna. Tento výsledok neslúži ako nezávislé hodnotenie modelu, ale potvrdzuje správnosť dátovej pipeline a konzistentnosť mapovania tried. V ML module je FinBERT k dispozícii ako záložný zdroj sentimentu v prípade, že vlastný model nie je nakonfigurovaný.

4.2.4 Kvalitatívna inferencia

Na ilustráciu správania boli zostavené tri krátke texty zodpovedajúce doménovému štýlu PhraseBanku (finančné faktografické vety s číslami a odbornými výrazmi). Každý text bol predložený všetkým trom modelom. Tabuľka 4.3 uvádza predikovanú triedu, pričom pre LR a LSTM sú v zátvorkách pravdepodobnosti tried (neg / neu / pos), pri FinBERT skóre $s \in [-1, 1]$ a pravdepodobnosti v tom istom poradí.

Tabuľka 4.3: Inferencia sentimentu na troch vzorových textoch.

Text (skrátene)	LR	LSTM	FinBERT
<i>CloudCore Inc. beat analysts' expectations...net sales up 28 %...</i>	positive (0,144 / 0,045 / 0,811)	positive (0,000 / 0,000 / 1,000)	positive, $s=0,940$ (0,019 / 0,021 / 0,959)
<i>Horizon Bio's net sales fell...pre-tax loss...FDA rejected...</i>	negative (0,751 / 0,037 / 0,212)	negative (0,989 / 0,010 / 0,001)	negative, $s=-0,967$ (0,974 / 0,019 / 0,007)
<i>Midland Utilities...shareholder meeting...no change to dividend...</i>	neutral (0,061 / 0,844 / 0,094)	neutral (0,000 / 1,000 / 0,000)	neutral, $s=-0,052$ (0,080 / 0,893 / 0,027)

Všetky tri modely správne klasifikujú všetky tri texty. Pozoruhodným rozdielom je distribúcia pravdepodobností: LR produkuje rozložené odhady (napr. 0,751 pre triedu negative pri negatívnom texte), LSTM vykazuje extrémnu saturáciu softmax vrstvy (0,989)

a FinBERT dáva vysoké, ale menej saturované hodnoty vrátane nenulových odhadov pre vedľajšie triedy. Saturácia LSTM je typická pre neurónové modely bez kalibrácie — model môže priradiť vysokú mieru istoty aj vstupu mimo distribúcie tréningových dát [31].

Text mimo distribúcie tréningových dát. Pri dlhšom analytickom článku zo zdroja Yahoo Finance (článok o akcii SoundHound AI, 32 riadkov), ktorého dominantný tón je z finančného hľadiska negatívny, sa modely rozchádzajú:

- **LR:** predikcia **positive**.
Pravdepodobnosti neg / neu / pos = 0,340 / 0,173 / 0,487.
- **LSTM:** predikcia **neutral**.
Pravdepodobnosti neg / neu / pos = 0,001 / 0,999 / 0,000.
- **FinBERT:** predikcia **negative**. Skóre $s = -0,911$.
Pravdepodobnosti neg / neu / pos = 0,923 / 0,065 / 0,012.

Článok popisuje pokles ceny akcie o 65 % od 52-týždenného maxima, stratu a neistotu ohľadom ziskovosti, teda obsah, ktorý by pri celkovom čítaní mal viesť k negatívnemu sentimentu. FinBERT, ktorý spracúva celý vstup ako sekvenciu tokenov, s touto interpretáciou súhlasí. LR s TF-IDF n-gramami však rozpoznáva aj pozitívnu lexiku investičnej analýzy (*investable opportunity, buying opportunity*), čo tlačí výstup smerom k triede positive a odchýli ho od negatívneho charakteru textu. LSTM priradí triede neutral extrémnu saturovanú pravdepodobnosť (0,999). Tento príklad ilustruje, že modely trénované na krátkych izolovaných vetách PhraseBanku sa pri dlhých textoch mimo ich tréningovej distribúcie správajú odlišne, pričom FinBERT vykazuje vyššiu doménovú robustnosť voči celkovému negatívnemu tónu článku. Výsledky na tomto texte nie sú súčasťou kvantitatívneho hodnotenia, ale slúžia na ilustráciu limitov modelov pri nasadzovaní na produkčné správy.

4.2.5 Zhrnutie

Na testovacej množine PhraseBanku (Sentences_50Agree) dosahuje TF-IDF + LR makro-F1 = 0,763 a LSTM makro-F1 = 0,701. Napriek

tomu, že LSTM je architekturne komplexnejší model, pri datasete tejto veľkosti (3 400 tréningových viet) je výhodou LR jeho odolnosť voči preučeniu a efektívnosť pri riedkej TF-IDF reprezentácii. Oba výsledky sú dosiahnuté za identických podmienok rozdelenia dát, čím je zabezpečená porovnateľnosť.

4.3 Vyhodnotenie modelov predikcie cenového pohybu

Táto podkapitola prezentuje výsledky experimentov s tromi architekturami pre predikciu denného pohybu ceny — logistickou regresiou (LR), gradientovým zosilňovaním LightGBM a rekurentnou sieťou LSTM. Pre každú architektúru boli vyhodnotené až tri varianty:

1. **Jednoakciový model** — tréning výlučne na príznakoch akcie AAPL (cenové príznaky vrátane technických indikátorov RSI a MACD).
2. **Združený model (MAG7)** — tréning na dátach siedmich titulov (AAPL, MSFT, GOOGL, META, NVDA, TSLA, AMZN). Časový rez sa počíta pre každý symbol samostatne a množiny sa spoja. Tento variant nie je v súčasnej implementácii podporovaný pre LSTM: po zreťazení (poolingu) riadkov viacerých titulov by niektoré vstupné okná prekryvali hranicu medzi dvoma rôznymi akciami, čo by viedlo k nezmyselným sekvenciám.
3. **Jednoakciový model s príznakmi sentimentu** — tréning na AAPL s rozšírenými príznakmi z historických správ (polarity sentimentu, kategórie udalostí, pozornosť médií a pod.).

Cieľová premenná nadobúda dve triedy podľa znamienka jednodňového výnosu r : UP pri $r \geq 0$ a DOWN pri $r < 0$. Symbol r označuje relatívnu zmenu uzávierkovej ceny medzi dvoma po sebe idúcimi obchodnými dňami. Dni s $|r| \leq 0,75\%$ sú v dátach označené ako „v pásme“ (in_band) a ešte pred jeho časovým rozdelením na tréningovú, validačnú a testovú množinu sa z neho vynechajú, aby model nemusel rozlišovať šumové pohyby v blízkosti nulového výnosu. Príznakový priestor zahŕňa 41 cenových stĺpcov (výnosy, momenta, vola-

tilitu, kľzavé priemery, objemové pomery, RSI a MACD). Sentimentový variant pridáva 14 ďalších stĺpcov.

4.3.1 Kvantitatívne výsledky

Tabuľka 4.4 zhŕňa výsledky na validačnej a testovacej množine pre osem natrénovaných modelov a dva varianty momentum baseline. Najlepšia hodnota v každom stĺpci tabuľky je zvýraznená tučným písmom.

Tabuľka 4.4: Porovnanie modelov predikcie cenového pohybu.

Model	Variant	Val bal. acc.	Val makro-F1	Test bal. acc.	Test makro-F1
Momentum baseline	AAPL	0,492	0,492	0,516	0,516
	MAG7	0,492	0,492	0,536	0,536
LR	AAPL	0,501	0,442	0,484	0,385
	MAG7	0,504	0,487	0,506	0,462
	AAPL + sent.	0,520	0,480	0,506	0,387
LightGBM	AAPL	0,548	0,539	0,562	0,553
	MAG7	0,524	0,516	0,521	0,514
	AAPL + sent.	0,536	0,477	0,512	0,428
LSTM	AAPL	0,504	0,467	0,497	0,494
	AAPL + sent.	0,524	0,522	0,494	0,410

Tabuľka obsahuje aj momentum baseline, ktorý predikuje, že zajtra cena pôjde rovnakým smerom ako dnes (UP pri kladnom výnose, DOWN pri zápornom). Na testovacej množine dosahuje vyváženú presnosť 0,516 pre AAPL a 0,536 pre MAG7. Prekonáva ho iba LightGBM trénovaný výhradne na príznakoch AAPL (0,562), zatiaľ čo ostatné modely za ním zaostávajú. Celkový výkon je výrazne nižší než pri sentimentovej klasifikácii.

4.3.2 Jednoakciový vs. združený tréning

Združený tréning (MAG7) prináša odlišné výsledky pre LR a LightGBM:

- **LR:** Zväčšenie trénovacej množiny z $\sim 1\,200$ na $\sim 8\,400$ pozorovaní zlepšuje testovú vyváženú presnosť o 2,2 percentuálneho bodu a makro-F1 o 7,7 bodu. Logistická regresia profituje z väčšieho objemu dát, pretože má malú kapacitu a stabilnejší odhad koeficientov pri vyššom n .

- **LightGBM:** Združený model naopak testovú vyváženú presnosť znižuje o 4,1 bodu. Stromové modely sú schopné zachytiť vzory špecifické pre jednu akciu (napr. vzťah medzi RSI a konkrétnym stupňom volatility AAPL). Pridanie šiestich ďalších titulov so štatisticky odlišnými distribúciami príznakov tieto vzory rozrieduje, čo vedie k horšej generalizácii na jednoakciovom teste.

LightGBM vykazuje najvýraznejší rozdiel medzi tréningovým a testovým výkonom: tréningová vyvážená presnosť jednoakciového modelu dosahuje 0,832, zatiaľ čo testová len 0,562. Pokles o 27 percentuálnych bodov poukazuje na preučenie napriek tomu, že early stopping bol aktívny, pričom najlepší výsledok na validačnej množine bol pri early stoppingu dosiahnutý už v 5. kole. Združený model mierne tlmí preučenie (tréning 0,654 vs. test 0,521, pokles 13,3 b.).

4.3.3 Vplyv sentimentových príznakov

Pridanie 14 príznakov odvodených zo správ (sentimentová polarita, kategórie udalostí, mediálna pozornosť) neviedlo k jednoznačnému zlepšeniu:

- **LR:** Vyvážená presnosť na teste stúpala o 2,2 b. (z 0,484 na 0,506), no makro-F1 sa takmer nezmenilo (0,385 vs. 0,387). Logistická regresia teda čiastočne využije priemerný sentiment, ale nedokáže ho skombinovať s cenovými príznakmi spôsobom, ktorý by zlepšil prediktívny výkon naprieč oboma triedami.
- **LightGBM:** Oba testovacie ukazovatele klesli — vyvážená presnosť o 5,0 b. (z 0,562 na 0,512), makro-F1 o 12,5 b. (z 0,553 na 0,428). Analýza dôležitosti príznakov (tabuľka 4.5) ukazuje, že väčšina sentimentových stĺpcov získala nulovú dôležitosť — vrátane všetkých príznakov kategórií udalostí a mediálnej pozornosti. Nenulový príspevok mali len stĺpce priemerného negatívneho a neutrálneho sentimentu. Pridanie dimenzií bez signálu zvyšuje riziko šumu a sťažuje regularizáciu.
- **LSTM:** Na validácii sentimentový variant mierne zlepšil vyváženú presnosť (0,524 vs. 0,504), ale na testovacej množine klesol na 0,494. Rozšírenie vstupnej dimenzie (z 41 na 55 príznakov)

bez zodpovedajúceho nárastu trénovacích dát vedie k nestabilnej generalizácii.

Sentimentové príznaky pokrývajú správy o AAPL za roky 2020 až 2024 (celkovo $\sim 1\,530$ článkov). Dni bez správ sa vo všetkých sentimentových stĺpcoch doplnia hodnotou 0, takže trénovacie okno zostáva rovnaké ako pri modeli bez sentimentu. Avšak prvých zhruba päť rokov cenových dát (2015 – 2019) obsahuje pre všetky sentimentové stĺpce iba nuly — model sa teda v tejto časti trénovacej množiny nemôže opierať o žiadny sentimentový signál a nulové hodnoty interferujú so skutočnými hodnotami z neskorších rokov. Okrem toho je hustota správ nerovnomerná: rok 2020 obsahuje len 121 článkov oproti ~ 360 článkov v rokoch 2022 až 2024.

Na overenie tohto predpokladu bol pre každú architektúru spustený doplnkový beh s ohraničením dát na roky 2020 – 2024, teda na obdobie s pravidelným denným spravodajstvom. Pri logistickej regresii toto ohraničenie prinieslo výrazné zlepšenie: testové makro-F1 sentimentového variantu stúplo z 0,387 na 0,506, zatiaľ čo vyvážená presnosť zostala na 0,506. Čistejší sentimentový signál bez šumu núl teda pomohol modelu logistickej regresii. Pre LightGBM a LSTM ohraničenie zmenšilo trénovaciu množinu z $\sim 2\,500$ na $\sim 1\,000$ pozorovaní, čo oba modely poškodilo: testová vyvážená presnosť LightGBM klesla z 0,562 na 0,452 a LSTM z 0,497 na 0,483. Tieto modely sú citlivejšie na objem trénovacích dát než na čistotu sentimentového signálu.

4.3.4 Dôležitosť príznakov (LightGBM)

LightGBM umožňuje priamu kvantifikáciu dôležitosti príznakov prostredníctvom počtu rozdelení v stromoch (*split count*). Tabuľka 4.5 uvádza desať najdôležitejších príznakov pre jednoakciový model na AAPL bez sentimentových stĺpcov.

Tabuľka 4.5: Desať príznakov podľa počtu rozdelení v stromoch.

Príznak	Počet rozdelení
return_1d	13
return_lag_3	11
price_vs_ma_20	9
volatility_20d	8
macd_hist	8
return_mean_10d	7
return_10d	6
volatility_5d	6
volume_change_lag_1	6
volume_ratio_5	6

Dominujú krátkodobé výnosové príznaky (jednodenný výnos a oneskorený výnos z 3 dní), dlhodobé technické indikátory (pomer ceny k 20-dňovému kľzavému priemeru a MACD histogram) a miera volatility (20-dňová). Naopak štandardné odchýlky výnosov a pomer krátkodobých kľzavých priemerov nezískali ani jedno rozdelenie, čo naznačuje redundanciu s inými stĺpcami.

4.3.5 Diskusia

Predikcia denného pohybu ceny je v odbornej literatúre považovaná za úlohu na hranici efektívneho trhu [3]. Výsledky experimentov sú konzistentné s touto teóriou — aj najlepší model (LightGBM, AAPL) prevýšil momentum baseline len o 4,6 percentuálneho bodu vyváženej presnosti (0,562 vs. 0,516). Na porovnanie, sentimentové modely dosahovali výrazne vyššie makro-F1 (viď tabuľku 4.1), čo je kvalitatívne odlišná úroveň úspešnosti.

Niekoľko faktorov obmedzuje výkon cenových modelov:

1. **Efektívnosť trhu.** Verejne dostupné cenové príznaky sú už do značnej miery zarátané do aktuálnej ceny. Stromový model dokáže zachytiť nelineárne kombinácie príznakov (napr. súčasný výskyt nízkej volatility a vysokého momentu), no zisk je marginálny.

2. **Časový posun distribúcie.** Rozdelenie je časové — testová množina pokrýva iné trhové podmienky než trénovacia. Preučenie LightGBM (tréning 0,832 vs. test 0,562) ilustruje, ako stromový model zachytáva vzory špecifické pre historické obdobie, ktoré sa v testovacom období nereprodukuje.
3. **Riedkosť sentimentového signálu.** Správy o AAPL pokrývajú roky 2020 až 2024, pričom dni bez správ sú vyplnené nulou. Pre prvých päť rokov cenových dát (2015 – 2019) sú všetky sentimentové stĺpce nulové, čo stiera rozdiel medzi skutočnou absenciou správ a neutrálnym dňom s aktívnym spravodajstvom. Model tak nedokáže konzistentne využiť sentimentový signál naprieč celým trénovacím radom.
4. **Výber prahu pre pásmo šumu.** Prah 0,75 % určuje dni v pásme, ktoré sa pred časovým rozdelením vynechajú. Vyšší prah by zmenšil počet pozorovaní a zanechal len dni s výraznejším pohybom, nižší prah by naopak zahrnul viac šumových dní a sťažil modelu nájsť konzistentný signál.

4.3.6 Zhrnutie

LightGBM trénovaný na jednoakciových príznakoch AAPL dosahuje najlepší testový výkon (vyvážená presnosť 0,562, makro-F1 0,553). Združený tréning na MAG7 zlepšuje logistickú regresiu, ale zhoršuje LightGBM. Sentimentové príznaky z krátkych spravodajských radov neprinášajú štatisticky jednoznačné zlepšenie — väčšina sentimentových stĺpcov nadobúda v LightGBM nulovú dôležitosť. LSTM na denných dátach dosahuje výkon blízky náhodnej predikcii, čo súvisí s obmedzeným počtom trénovacích sekvencií a absenciou dlhodobých závislostí v jednodennom horizonte. Celkovo výsledky potvrdzujú, že predikcia denného pohybu ceny predstavuje podstatne ťažšiu úlohu než sentimentová klasifikácia, a výsledky sú konzistentné s hypotézou efektívneho trhu.

4.4 Praktická relevantnosť a limity modelov

Táto podkapitola zasadzuje výsledky experimentov do širšieho kontextu. Cieľom nie je opakovať zistenia uvedené v podkapitolách 4.2 a 4.3, ale posúdiť ich praktický význam a hranice.

4.4.1 Čo z metrík plynie pre rozhodovanie

Sentimentové modely dosahujú na datasete Financial PhraseBank solídne výsledky (viď tabuľku 4.1). Tento výsledok však platí pre krátke, faktograficky formulované vety, ktoré tvorili trénovacie dáta. Produkčné správy sú typicky dlhšie, štruktúrnejšie a obsahujú zmiešané signály. Príklad analýzy zo zdroja Yahoo Finance v podkapitole 4.2 ukazuje, že modely trénované na PhraseBanku sa na takýchto textoch rozchádzajú. Vysoká testová presnosť teda neopodstatňuje automatické rozhodnutia postavené výhradne na výstupe sentimentového klasifikátora.

Pri predikcii cenového pohybu je obmedzenie ešte výraznejšie: najlepší model (LightGBM na AAPL) prevýšil momentum baseline len o 4,6 percentuálneho bodu vyváženej presnosti (0,562 vs. 0,516). V praxi však takéto prevýšenie nestačí na ziskovú stratégiu: predikcia je denná, transakčné náklady, sklznutie ceny pri realizácii a riadenie rizika znižujú reálnu hodnotu marginálnej informácie. Výstupy modelu je preto vhodné chápať ako doplnkový signál, nie ako samostatný základ pre obchodné rozhodnutia.

4.4.2 Limity dát a výberu vzorky

Experimenty s predikciou cien sú primárne postavené na jednej akcii (AAPL) a doplnkovo na MAG7 — siedmich amerických technologických tituloch s podobným trhovým profilom. Generalizácia na iné triedy aktív (napr. malé spoločnosti, komodity, dlhopisy) nie je overená a nie je z výsledkov odvoditeľná.

Dáta historických správ o AAPL pokrývajú roky 2020 – 2024, zatiaľ čo cenové rady siahajú desať rokov dozadu. Pre roky 2015 – 2019 sú všetky sentimentové stĺpce nulové, čo obmedzuje využiteľnosť sentimentu ako príznaku. Doplnkový experiment s ohraničením na roky 2020 – 2024 potvrdil, že čistejší sentimentový signál pomáha

logistickej regresii, no stromové modely a LSTM pri menšom datasete strácajú.

4.4.3 Možné skreslenia

Skreslenie cieľovej premennej. Triedy UP a DOWN sú definované znamenkom jednodňového výnosu r , pričom dni s $|r| \leq 0,75\%$ sa pred časovým rozdelením vynechávajú ako šumové. Tento prah zodpovedá nastaveniu produkčnej služby, nie nevyhnutne ekonomicky zmysluplnej hranici. Iný prah by zmenil počet tréningových pozorovaní aj distribúciu tried. Modely sa učia na zvolenej definícii pojmu „pohyb ceny“ v dátach, nie na „reálnom zisku investora“ — rozdiel, ktorý pri interpretácii výsledkov treba mať na pamäti.

Časové skreslenie. Tréningové dáta pokrývajú väčšinou rastový trhový režim (2015 – 2024 s výnimkou korekcií). Testová množina je chronologicky posledných 15 % radu, teda pokrýva iné trhové podmienky než tréning, čo sa prejavuje poklesom výkonu pri nasadení modelu na nové dáta.

Skreslenie pozornosti médií. Hustota správ nie je rovnomerná: v dňoch s výsledkami hospodárenia alebo produktovými udalosťami je správ výrazne viac. Sentimentový signál je teda silnejší práve v dňoch so zvýšenou volatilitou ceny, čo môže vytvárať zdanlivú koreláciu medzi sentimentom a pohybom ceny, ktorá je artefaktom selektívneho spravodajstva, nie kauzálneho vzťahu.

Saturácia pravdepodobností. LSTM klasifikátor sentimentu vykazuje extrémnu saturáciu softmax vrstvy (pravdepodobnosť dominantnej triedy blízko 1,000). Bez kalibrácie nie sú tieto hodnoty interpretovateľné ako skutočné pravdepodobnosti udalosti — model môže priradiť vysokú istotu aj vstupu výrazne odlišnému od tréningových dát.

4.4.4 Interpretácia výstupov v kontexte aplikácie

Implementované modely slúžia ako orientačné analytické nástroje. Sentimentový klasifikátor poskytuje rýchlu kategorizáciu tónu správy,

ktorá môže byť užitočná pri spracovaní väčšieho objemu textu. Výstup modelu predikcie pohybu ceny indikuje pravdepodobnostnú tendenciu pohybu, nie záväzný signál. Oba modely sú experimentálneho charakteru a nenahrádzajú fundamentálnu ani technickú analýzu, riadenie rizika ani odborný úsudok.

Záver

Cieľom tejto práce bolo navrhnuť, implementovať a experimentálne overiť analytickú platformu FinPulse pre spracovanie finančných správ a predikciu pohybu cien akcií. FinPulse integruje modul strojového učenia s plne funkčnou webovou aplikáciou dostupnou bežnému používateľovi, čím premostuje medzeru medzi metodologicky hodnotnými, no finančne nedostupnými profesionálnymi nástrojmi a používateľsky príjemnými, no analyticky obmedzenými voľne dostupnými riešeniami.

Kapitola 1 vybuďovala teoretický základ práce. Zhrnutím výskumu v oblasti hypotézy efektívnych trhov, behaviorálnych financií, evolúcie NLP metód od Bag-of-Words cez slovné vnorenia až po transformové modely a architektúry hlbokého učenia pre finančné časové rady bolo zdokumentované, že priestor pre nenáhodné predikcie existuje, no je obmedzený a nestabilný. Teoretické závery priamo motivujú hybridný prístup kombinujúci cenové a textové vstupy.

Kapitoly 2 a 3 sa venovali návrhu a implementácii FinPulse. Architektúra je rozdelená do piatich vrstiev — od zberu dát cez analýzu správ a predikciu ceny až po serverovú logiku a klientskú aplikáciu. Návrh dôsledne oddeľuje tréningovú fázu od produkčnej inferencie: produkčný model predikcie pohybu ceny je postavený výhradne na cenových príznakoch, pričom experimenty so sentimentovými príznakmi boli vykonané ako oddelené tréningové behy. Implementácia zahŕňa konštrukciu a tréning všetkých modelov s podporou Azure Machine Learning pre výpočtovo náročné tréningové behy.

Kapitola 4 prezentovala výsledky experimentov. V úlohe sentimentovej klasifikácie na datasete Financial PhraseBank dosiahol model TF-IDF + LR makro-F1 = 0,763 a model LSTM makro-F1 = 0,701, pričom oba výsledky boli dosiahnuté za identických podmienok a preukázali stabilnú generalizáciu. V úlohe predikcie denného pohybu ceny najlepší výkon dosiahol model LightGBM tréňovaný na jednoakciových príznakoch AAPL (vyvážená presnosť = 0,562, makro-F1 = 0,553), čo predstavuje prevýšenie momentum baseline o 4,6 percentuálneho bodu. Tieto výsledky sú konzistentné s hypotézou efektívneho trhu — denná predikcia pohybu ceny zostáva podstatne ťažšou úlohou než sentimentová klasifikácia. Pridanie sentimentových príznakov k ceno-

vým neviedlo k jednoznačnému zlepšeniu: logistická regresia z nich čiastočne profitovala, kým LightGBM a LSTM skôr trpeli zníženou kvalitou dát v rokoch bez pravidelného spravodajstva.

Na základe uvedených zistení možno formulovať niekoľko smerov pre ďalší výskum. Kalibrácia pravdepodobností sentimentových modelov (napr. temperature scaling) by zvýšila spoľahlivosť ich výstupov ako vstupov do rozhodovacích systémov. Rozšírenie sentimentového datasetu o reálne produkčné správy v kratšom, homogénnejšom časovom okne by priamo adresovalo problém riedkosti signálu, ktorý oslaboval hybridné modely. Pre predikciu pohybu ceny by testovanie na širšej vzorke titulov z rôznych sektorov overilo mieru generalizovateľnosti zistení nad rámec technologického sektoru. Napokon, integrácia explicitnejšej správy rizika do výstupov predikčného modulu, napríklad odhadov neistoty, by zvýšila jeho praktickú využiteľnosť ako orientačného analytického nástroja.

Výsledná platforma FinPulse prekladá výskumné metódy NLP a strojového učenia do formy prístupnej individuálnemu investorovi. Zahrnuté modely sú experimentálneho charakteru a nenahrádzajú odborný úsudok, fundamentálnu analýzu ani riadenie rizika. Práca však dokumentuje, že aj relatívne jednoduché modely trénované na verejne dostupných dátach dokážu extrahovať merateľný signál, a že transparentné vykazovanie ich limitov je rovnako hodnotným výstupom ako samotné čísla.

Bibliografia

1. HUANG, Allen H.; WANG, Hui; YANG, Yi. FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research*. 2023, vol. 40, no. 2, s. 806–841. Dostupné z [doi: 10.1111/1911-3846.12832](https://doi.org/10.1111/1911-3846.12832).
2. ASHTIANI, Matin N.; RAAHEMI, Bijan. News-based Intelligent Prediction of Financial Markets Using Text Mining and Machine Learning: A Systematic Literature Review. *Expert Systems With Applications*. 2023, vol. 217, s. 119509. Dostupné z [doi: 10.1016/j.eswa.2023.119509](https://doi.org/10.1016/j.eswa.2023.119509).
3. DU, Kelvin; XING, Frank; MAO, Rui; CAMBRIA, Erik. Financial Sentiment Analysis: Techniques and Applications. *ACM Computing Surveys*. 2024, vol. 56, no. 9, s. 220. Dostupné z [doi: 10.1145/3649451](https://doi.org/10.1145/3649451).
4. CHIONG, Raymond; FAN, Zongwen; HU, Zhongyi; DHAKAL, Sandeep. A Novel Ensemble Learning Approach for Stock Market Prediction Based on Sentiment Analysis and the Sliding Window Method. *IEEE Transactions on Computational Social Systems*. 2022, vol. 10, no. 5, s. 2613–2623. Dostupné z [doi: 10.1109/TCSS.2022.3182375](https://doi.org/10.1109/TCSS.2022.3182375).
5. FAMA, Eugene F. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*. 1970, roč. 25, č. 2, s. 383–417. Dostupné z [doi: 10.1111/j.1540-6261.1970.tb00518.x](https://doi.org/10.1111/j.1540-6261.1970.tb00518.x).
6. AGRAWAL, Manish; SHUKLA, Piyush Kumar; NAIR, Rajit; NAYYAR, Anand; MASUD, Mehedi. Stock Prediction Based on Technical Indicators Using Deep Learning Model. *Computers, Materials & Continua*. 2022, vol. 70, no. 1, s. 287–304. Dostupné z [doi: 10.32604/cmc.2022.014637](https://doi.org/10.32604/cmc.2022.014637).
7. ZHAO, Yanli; YANG, Guang. Deep Learning-based Integrated Framework for Stock Price Movement Prediction. *Applied Soft Computing*. 2023, vol. 133, s. 109921. Dostupné z [doi: 10.1016/j.asoc.2022.109921](https://doi.org/10.1016/j.asoc.2022.109921).

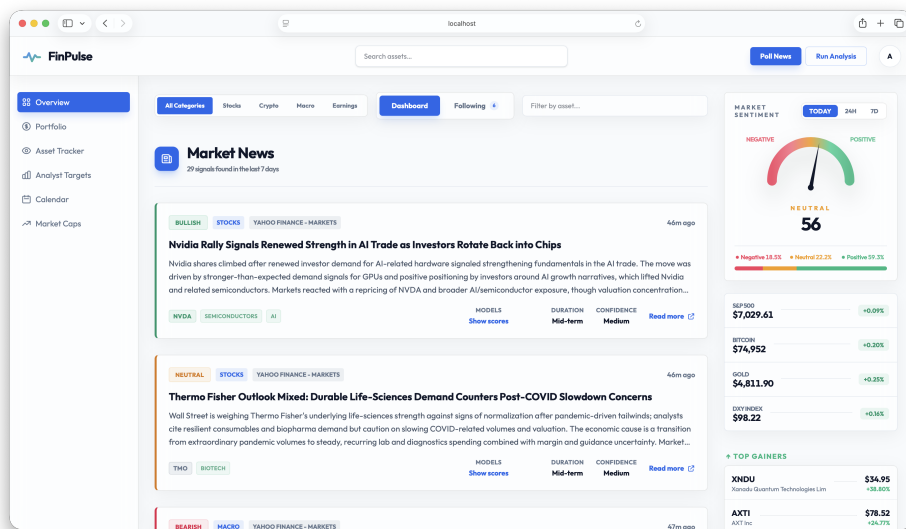
8. SALTON, Gerard; WONG, Anita; YANG, Chung-Shu. A Vector Space Model for Automatic Indexing. *Communications of the ACM*. 1975, roč. 18, č. 11, s. 613–620. Dostupné z [doi: 10.1145/361219.361220](https://doi.org/10.1145/361219.361220).
9. JONES, Karen Spärck. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*. 1972, roč. 28, č. 1, s. 11–21. Dostupné z [doi: 10.1108/eb026526](https://doi.org/10.1108/eb026526).
10. MIKOLOV, Tomas; CHEN, Kai; CORRADO, Greg; DEAN, Jeffrey. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*. 2013. Dostupné tiež z: <https://arxiv.org/abs/1301.3781>.
11. PENNINGTON, Jeffrey; SOCHER, Richard; MANNING, Christopher D. GloVe: Global Vectors for Word Representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014, s. 1532–1543. Dostupné z [doi: 10.3115/v1/D14-1162](https://doi.org/10.3115/v1/D14-1162).
12. VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Łukasz; POLOSUKHIN, Illia. Attention Is All You Need. In: *Advances in Neural Information Processing Systems*. 2017, zv. 30. Dostupné z [doi: 10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762).
13. DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton; TOUTANOVA, Kristina. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of NAACL-HLT*. 2019, s. 4171–4186. Dostupné z [doi: 10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423).
14. ARACI, Dogu. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. *arXiv preprint arXiv:1908.10063*. 2019. Dostupné z [doi: 10.48550/arXiv.1908.10063](https://doi.org/10.48550/arXiv.1908.10063).
15. LOUGHRAN, Tim; MCDONALD, Bill. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*. 2011, roč. 66, č. 1, s. 35–65. Dostupné z [doi: 10.1111/j.1540-6261.2010.01625.x](https://doi.org/10.1111/j.1540-6261.2010.01625.x).

16. RADFORD, Alec; NARASIMHAN, Karthik; SALIMANS, Tim; SUTSKEVER, Ilya. *Improving Language Understanding by Generative Pre-Training* [Technical report, OpenAI]. 2018. Dostupné tiež z: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
17. FJELLSTRÖM, Carmina. *Long Short-Term Memory Neural Network for Financial Time Series*. 2022. Dostupné z arXiv: 2201.08218.
18. SONKAVDE, Gaurang; DHARRAO, Deepak Sudhakar; BONGALE, Anupkumar M.; DEOKATE, Sarika T.; DORESWAMY, Deepak; BHAT, Subraya Krishna. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies*. 2023, vol. 11, no. 3, s. 94. Dostupné z doi: 10.3390/ijfs11030094.
19. TRADING 212. *Trading 212 API Documentation* [<https://docs.trading212.com/api>]. 2026. Prístup: 10.05.2026.
20. PYTHON SOFTWARE FOUNDATION. *Python Documentation* [<https://docs.python.org/3/>]. 2026. Prístup: 10.05.2026.
21. FASTAPI. *FastAPI Documentation* [<https://fastapi.tiangolo.com/>]. 2026. Prístup: 10.05.2026.
22. SQLITE. *SQLite Documentation* [<https://www.sqlite.org/docs.html>]. 2026. Prístup: 10.05.2026.
23. SQLALCHEMY. *SQLAlchemy Documentation* [<https://docs.sqlalchemy.org/>]. 2026. Prístup: 10.05.2026.
24. REACT. *React Documentation* [<https://react.dev/>]. 2026. Prístup: 10.05.2026.
25. TYPESCRIPT. *TypeScript Documentation* [<https://www.typescriptlang.org/docs/>]. 2026. Prístup: 10.05.2026.
26. OPENAI. *OpenAI API Documentation* [<https://platform.openai.com/docs>]. 2026. Prístup: 10.05.2026.
27. EARNINGSAPI. *EarningsAPI Documentation* [<https://www.earningsapi.com/docs>]. 2026. Prístup: 10.05.2026.

-
28. MICROSOFT. *Azure Machine Learning documentation* [<https://learn.microsoft.com/azure/machine-learning/>]. 2026. Prístup: 10.05.2026.
 29. OSSAI, Frank. *Apple Stock (AAPL): Historical Financial News Data* [Kaggle dataset]. 2024. Dostupné tiež z: <https://www.kaggle.com/datasets/frankossai/apple-stock-aapl-historical-financial-news-data>. Prístup: 10.05.2026.
 30. MALO, Pekka et al. *Financial PhraseBank* [https://huggingface.co/datasets/takala/financial_phrasebank]. 2014. Verzia na Hugging Face; prístup: 10.05.2026.
 31. GUO, Chuan; PLEISS, Geoff; SUN, Yu; WEINBERGER, Kilian Q. On Calibration of Modern Neural Networks. In: *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2017, s. 1321–1330. Dostupné tiež z: <https://proceedings.mlr.press/v70/guo17a.html>.

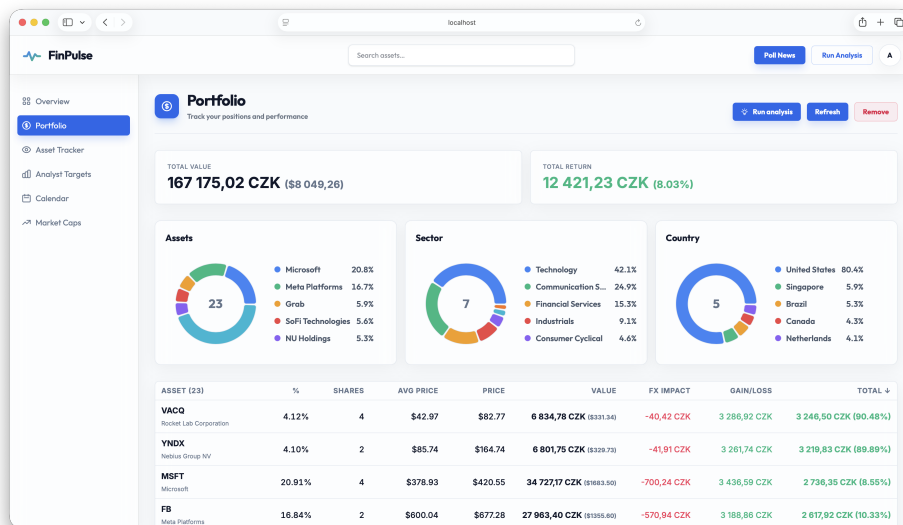
A Snímky používateľského rozhrania

Táto príloha dokumentuje grafické rozhranie FinPulse snímkami hlavných obrazoviek. Obrázky zodpovedajú stavu aplikácie pri vypracovaní práce a dopĺňajú stručný návrh v kapitole 2 a popis implementácie v kapitole 3.

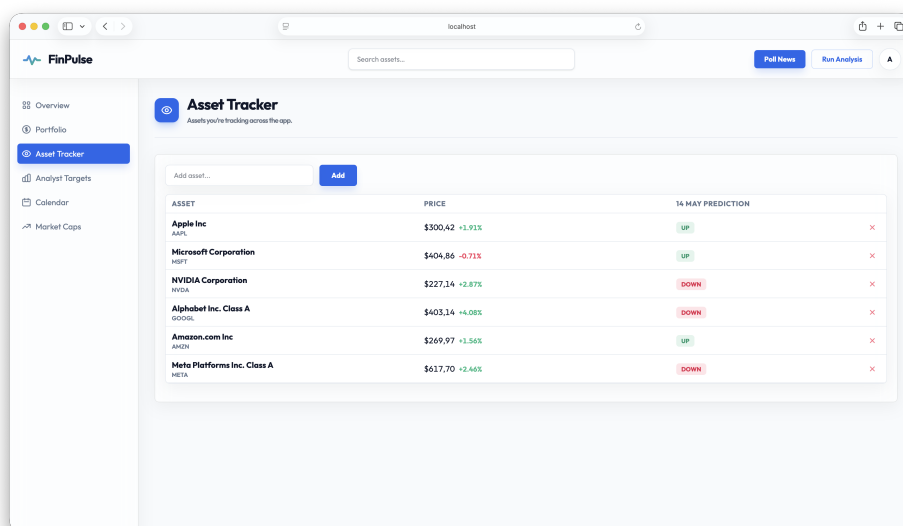


Obr. A.1: Prehľadová obrazovka *Overview* (dashboard, správy, ukazovatele).

A. SNÍMKY POUŽÍVATELSKÉHO ROZHRAVIA

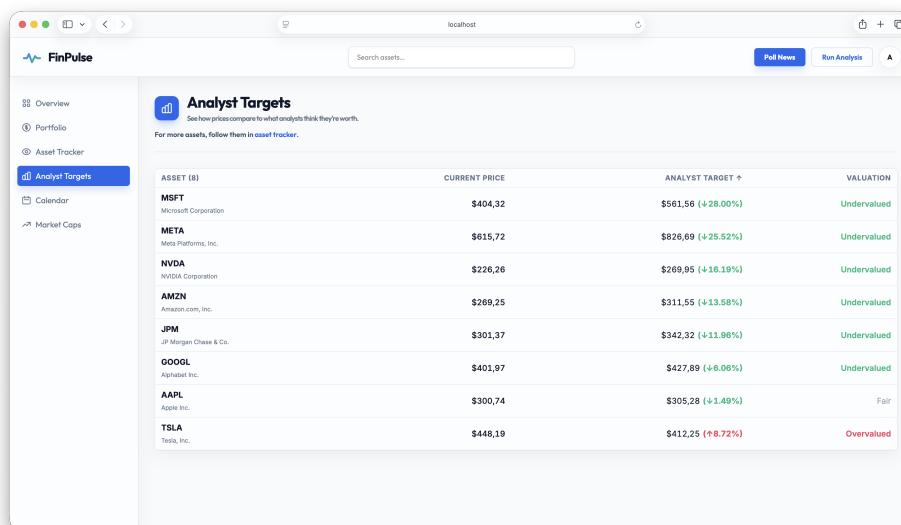


Obr. A.2: Obrazovka *Portfolio* po synchronizácii s brokerom.



Obr. A.3: Obrazovka *Asset Tracker* (sledované tituly).

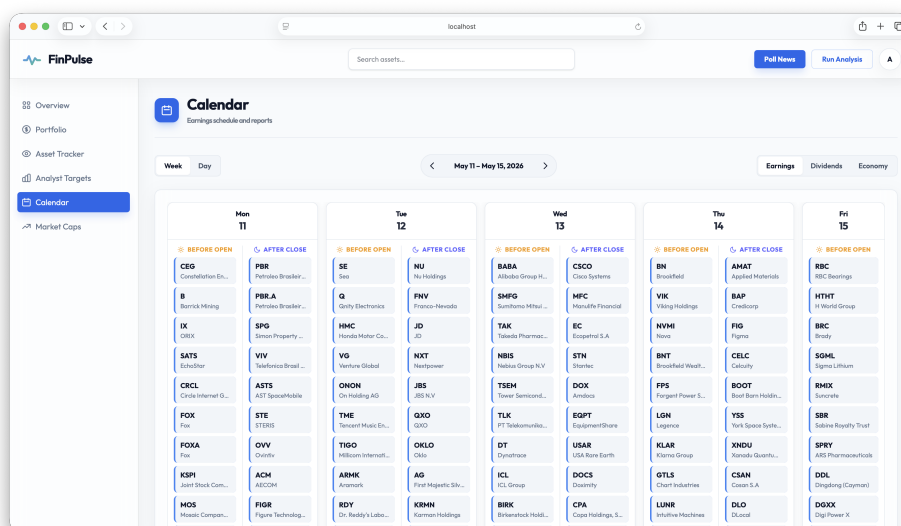
A. SNÍMKY POUŽÍVATELSKÉHO ROZHRAVIA



Analyst Targets
See how prices compare to what analysts think they're worth.
For more assets, follow them in [asset tracker](#).

ASSET (8)	CURRENT PRICE	ANALYST TARGET +	VALUATION
MSFT Microsoft Corporation	\$404,32	\$561,56 (+28.00%)	Undervalued
META Meta Platforms, Inc.	\$615,72	\$826,69 (+25.52%)	Undervalued
NVDA NVIDIA Corporation	\$226,26	\$269,95 (+16.19%)	Undervalued
AMZN Amazon.com, Inc.	\$289,25	\$311,55 (+7.58%)	Undervalued
JPM JP Morgan Chase & Co.	\$301,37	\$342,32 (+13.96%)	Undervalued
GOOGL Alphabet Inc.	\$401,97	\$427,89 (+6.06%)	Undervalued
AAPL Apple Inc.	\$300,74	\$305,28 (+1.49%)	Fair
TSLA Tesla, Inc.	\$448,19	\$412,25 (-7.72%)	Overvalued

Obr. A.4: Obrazovka *Analyst Targets* (index cieľových cien).



Calendar
Earnings schedule and reports

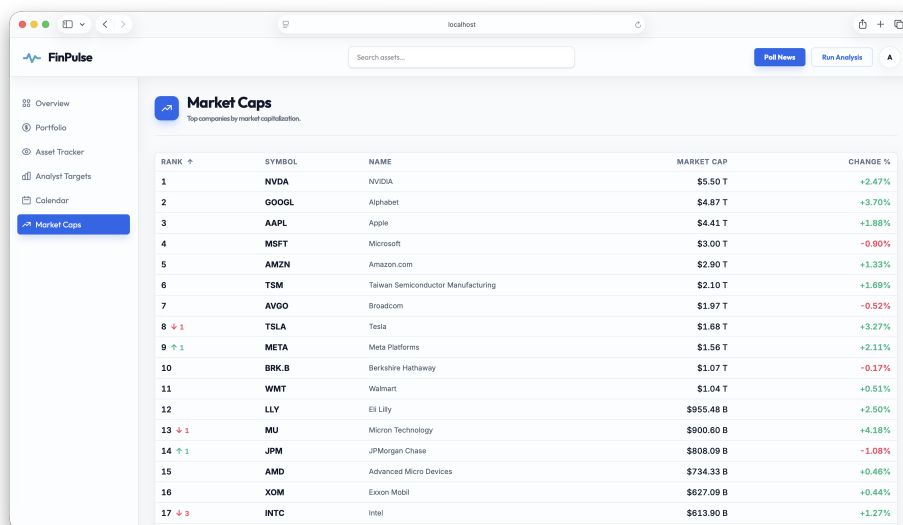
Week Day: May 11 - May 15, 2026

Earnings Dividends Economy

Mon 11		Tue 12		Wed 13		Thu 14		Fri 15	
BEFORE OPEN	AFTER CLOSE	BEFORE OPEN	AFTER CLOSE	BEFORE OPEN	AFTER CLOSE	BEFORE OPEN	AFTER CLOSE	BEFORE OPEN	AFTER CLOSE
CEG Covestro AG	PBR Permian Basin	SE Sea	NU Nu Holdings	BABA Alibaba Group	CSCO Cisco Systems	BN Brookfield	AMAT Applied Materials	RBC RBC Bearings	
B Barrick Mining	PBR.A Permian Basin	Q Qorvo	FNV Finmeccanica	SHW Shutterstock	MFC Monetta Financial	VIK Viking Holdings	BAP Bayer	HTHT Himalaya Group	
IK IQVIA	SPG Simon Property	JMC Jaguar Land Rover	JD JD.com	TAK Takeda Pharmaceutical	EC Ecoplast S.A.	NVPH Novartis	FIG Fidelity Investments	BRC Bristol-Myers Squibb	
SATS Satellite	VIV Telefonica Brasil	VG Venture Global	NXT Nextpower	NBS Nabors Group	STN Stantec	BNT Bristol-Myers Squibb	CELG Celgene	SGML Sigma Lithium	
CRCL Circle K	ASTS AST SpaceMobile	ONON On Holding AG	JBS JBS S.A.	TSEM Tower Semiconductor	DOX Amidex	FPS Fortinet	BOOT Boat International	RMX Rimac	
FOX Fox	STE STERIS	TME Tencent Music	QXO QXO	TLK PT Telekomunikacja	EGPT Egyptian	LGN Legence	YSS York Space Systems	SBR Sabine Royalty Trust	
FOXA Fox	OVV Ovintiv	TIGO Tigo Telecom	OKLO Oklo	DT Dynatrace	USAR USA Rare Earth	KLAR Klaro Group	XNDU Xenon Quantum	SPRY ASIS Pharmaceuticals	
KSPI Kaiser Steel	ACM AECOM	ARMK Armstrong	AG First Hygenic	ICL ICL Group	DOCS DocuSign	GTLN Gigamon	CSAN Cosan S.A.	DDL Dongling (Geyman)	
MOS Mosaic	FIOR Fior Technology	RDY Roche	KRHH Korhan Holdings	BIRX Biorix	CPA Copa Holdings	LUNR Lunar	DLO Data Domain	DOXX Dogecoin	

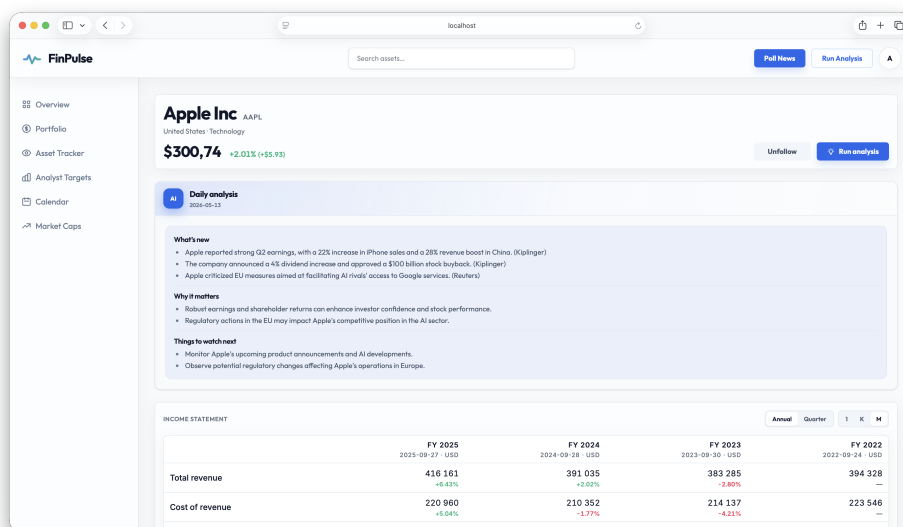
Obr. A.5: Obrazovka *Calendar* (výsledky, dividendy, makro).

A. SNÍMKY POUŽÍVATELSKÉHO ROZHRAVIA



RANK	SYMBOL	NAME	MARKET CAP	CHANGE %
1	NVDA	NVIDIA	\$5.50 T	+2.47%
2	GOOGL	Alphabet	\$4.87 T	+3.70%
3	AAPL	Apple	\$4.41 T	+1.88%
4	MSFT	Microsoft	\$3.00 T	-0.90%
5	AMZN	Amazon.com	\$2.90 T	+1.33%
6	TSM	Taiwan Semiconductor Manufacturing	\$2.10 T	+1.69%
7	AVGO	Broadcom	\$1.97 T	-0.52%
8	TSLA	Tesla	\$1.68 T	+3.27%
9	META	Meta Platforms	\$1.56 T	+2.11%
10	BRK.B	Berkshire Hathaway	\$1.07 T	-0.17%
11	WMT	Walmart	\$1.04 T	+0.51%
12	LLY	Eli Lilly	\$955.48 B	+2.50%
13	MU	Micron Technology	\$900.80 B	+4.18%
14	JPM	JPMorgan Chase	\$808.09 B	-1.08%
15	AMD	Advanced Micro Devices	\$734.33 B	+0.46%
16	XOM	Exxon Mobil	\$627.09 B	+0.44%
17	INTC	Intel	\$613.90 B	+1.27%

Obr. A.6: Obrázovka *Market Caps*.



	FY 2025	FY 2024	FY 2023	FY 2022
	2025-09-27 - USD	2024-09-28 - USD	2023-09-30 - USD	2022-09-24 - USD
Total revenue	416 161	391 035	383 285	394 328
Cost of revenue	220 960	210 352	214 137	223 546
	1.06 201	1.80 683	1.60 148	1.70 780

Obr. A.7: Detail konkrétneho titulu.

B Prompty pre OpenAI API

Táto príloha uvádza znenie promptov posielaných do OpenAI API. Zástupné symboly v tvare { . . . } sa za behu nahrádzajú skutočnými hodnotami.

B.1 Filter relevantnosti správy

Predbežné rozhodnutie o relevantnosti článku pred hĺbkovou analýzou. Do promptu vstupuje titulok a prvých 500 znakov tela článku. Očakávaná odpoveď je JSON.

System

```
You decide if a news article is relevant for financial market analysis. Reply only with the JSON.
```

User

```
Title: {article.title}
Content snippet: {article.raw_content[:500]}

Is this article worth analyzing for financial/market sentiment and impact? Consider: stocks, crypto, macro, earnings, trading, companies, economy. Skip: celebrity, lifestyle, sports, gossip, non-financial fluff.

Reply ONLY with JSON:
{ "worth_analyzing": true or false,
  "reason": "one short phrase" }
```

B.2 Analýza článku

Hĺbková analýza správy pre relevantné články. Telo článku je orezané na 2000 znakov. Očakávaná odpoveď je JSON.

System

You are a senior financial analyst providing high-signal market intelligence.

User

Analyze the following financial news article and provide a structured JSON response.

Title: {article.title}
 Content: {article.raw_content[:2000]}

Respond ONLY with a JSON object following this structure:

```
{
  "headline": "A bold, descriptive title summarizing the main event",
  "summary": "A 2-4 sentence concise summary focusing on causality (Event -> Economic Cause -> Market Reaction)",
  "sentiment": "Bullish, Bearish, or Neutral",
  "sentiment_score": float between -1.0 and 1.0,
  "assets": [
    { "symbol": "string",
      "impact": "positive" or "negative" or "neutral",
      "type": "ticker" or "sector" }
  ],
  "impact_duration": "Short-term, Mid-term, or Long-term",
  "confidence_score": "Low, Medium, or High",
  "category": "Stocks, Crypto, Macro, or Earnings"
}
```

For "assets": list of objects with "symbol", "impact", and "type".

- "symbol": only equity (stock) tickers, not ETFs; or short sector/theme labels (one or two words, uppercase). No long phrases or parentheses.
- "impact": "positive" if the news is bullish for that symbol, "negative" if bearish, "neutral" otherwise.
- "type": "ticker" for a listed stock symbol (e.g. a real ticker that trades on an exchange); "sector" for a sector or theme label (e.g. SEMICONDUCTORS, SAAS, AI, REITS, ENERGY). Use "ticker" only for actual stock tickers.

B.3 Investičný súhrn aktíva

Investičný súhrn aktíva zobrazovaný na stránke jeho detailu. Výstup je obyčajný text bez Markdownu.

System

```
You are a cautious senior financial analyst. Your task is to
summarize the most important recent developments, news, and
events related to the following asset that an investor should
know. Write plain text suitable for a simple product UI:
never use Markdown (no **, ##, or [title](url)).
```

User

```
Asset:
- ticker: {asset.ticker}
- name: {asset.name}

Instructions:
- Focus only on information that is important or interesting
  for investors
- Prioritize news that can impact price, sentiment, or
  long-term outlook
- Avoid generic explanations of what the asset is
- Use simple, clear language
- No hype, no emojis
- No financial advice

Output structure:
1. What's new (3-5 bullet points)
2. Why it matters (short explanation)
3. Things to watch next (1-3 points)

Keep it concise, insightful, and under 150 words.
```

B.4 Analýza portfólia

Analýza zloženia portfólia. Blok {portfolio_context} sa skladá dynamicky z aktuálnych dát. Očakávaná odpoveď je JSON.

System

```
You are a senior portfolio analyst. Your task is to analyze the portfolio and provide only advice and recommendations. Do not output a summary of the portfolio or a list of positions by ticker - the data is for context only. Output must contain only: strengths, weaknesses, risks, and diversification suggestions. Each item must be one clear sentence with concrete data or recommendation (percentages, facts). This is not personal financial advice; do not recommend specific buys or sells.
```

User

```
{portfolio_context}

---

Return the response ONLY as valid JSON with exactly these keys (all required):
- "overview": one short paragraph with overall advice/recommendation (not a recap of positions)
- "strengths": array of strings (portfolio strengths)
- "weaknesses": array of strings (portfolio weaknesses)
- "risks": array of strings (risks)
- "diversification_suggestions": array of strings (suggestions to improve diversification)

No other text outside the JSON.
```