

**SZÉCHENYI ISTVÁN EGYETEM
KAUTZ GYULA GAZDASÁGTUDOMÁNYI KAR**

**Európai olajipari társaságok negyedéves jelentéseinek
elemzése, a piacra gyakorolt hatásainak vizsgálata
természetes nyelvfeldolgozás használatával**

Készítette:

Szombathy Ádám

CJ4BI7

Gazdálkodási és menedzsment

Belső konzulens:
Dr. Tóth Árpád
egyetemi docens

2024

Tartalomjegyzék

FOGALOM MEGHATÁROZÁS.....	1
BEVEZETÉS	3
1. SPECIFIKUS INFORMÁCIÓ KÖZZÉTÉTELEK GAZDASÁGI HATÁSAINAK MÉRÉSE TERMÉSZETES NYELVFELDOLGOZÁS MÓDSZERÉVEL KIEGÉSZÍTVE	4
1.1. A TERMÉSZETES NYELVFELDOLGOZÁS SZAKIRODALMA	5
1.1.1. Természetes nyelvfeldolgozási módszerek és algoritmusok	6
1.1.2. Hangulatelemzés természetes nyelvfeldolgozással	6
1.1.3. Természetes nyelvfeldolgozás és gépi tanulás alkalmazása a tőzsdei elemzésekben.....	8
1.2. A TERMÉSZETES NYELVFELDOLGOZÁS LEHETSÉGES ALKALMAZÁSI TERÜLETEI	10
1.2.1. Részvénytőzsdén volatilitásának előrejelzése	10
1.2.2. Jegybankelnöki beszédek vizsgálata	11
1.2.3. A közvélemény és politikai helyzet hatása a tőzsdére.....	13
1.2.4. Közösségi média üzenetek és a tőzsdei árfolyamok kapcsolata	14
1.2.5. A természetes nyelvfeldolgozás jövője	15
2. TERMÉSZETES NYELVFELDOLGOZÁS MÓDSZERÉVEL PRIMER ÉS SZEKUNDER KUTATÁS EREDMÉNYEINEK BEMUTATÁSA.....	16
2.1. VÁLLALATOK BEMUTATÁSA ÉS SZEKUNDER KUTATÁS.....	16
2.1.1. Vállalatok bemutatása	17
2.1.2. Elemzett vállalatok legfontosabb pénzügyi adatai	18
2.2. PRIMER KUTATÁS	20
2.2.1. Negyedéves jelentések elemzése	20
2.2.2. Leíró statisztika	21
2.2.3. Szavak közötti korreláció vizsgálata	23
2.2.4. A részvény árfolyamok és a fenntarthatósághoz kötődő szavak száma közötti kapcsolat vizsgálata	33
2.3. KÖVETKEZTETÉSEK, JAVASLATOK	36
ÖSSZEGZÉS.....	38
ÁBRAJEGYZÉK	39
TÁBLÁZATJEGYZÉK	39
IRODALOMJEGYZÉK.....	41
MELLÉKLETEK	46

FOGALOM MEGHATÁROZÁS

- NLP:** A Natural Language Processing (magyarul: természetes nyelvfeldolgozás) egy olyan terület a mesterséges intelligencia (AI) és a számítógépes nyelvészet kereszteződésében, amely az emberi nyelvet dolgozza fel. Az NLP technológiák segítségével a számítógépek képesek értelmezni és válaszolni az emberi nyelven megfogalmazott kérdésekre vagy utasításokra.
- NLFF:** A Natural Language based Financial Forecasting (magyarul: természetes nyelv alapú pénzügyi előrejelzés) egy olyan módszer vagy rendszer, amely a természetes nyelv feldolgozás (NLP) technológiát alkalmazza a pénzügyi előrejelzések létrehozásához. Ez a megközelítés ötvözi a pénzügyi adatokat és az NLP-t azzal a céllal, hogy könnyebben megérthessük és értelmezhessek a pénzügyi információkat.
- BoW:** A BoW, vagyis "Bag of Words" (magyarul: szavak halmaza) egy egyszerű, de hatékony módszer a szövegek számítógépes feldolgozására és reprezentálására. A BoW lényege, hogy egy dokumentumot vagy szöveget egyetlen halmazba helyez, figyelmen kívül hagyva a szavak sorrendjét és struktúráját, csak a szavak előfordulását rögzítve. Tehát a dokumentumot egy vektorral reprezentálja, ahol minden dimenzió egy adott szó és a kapott érték a szó előfordulásának gyakoriságát vagy jelenlétét mutatja.
- TF-IDF:** A TF-IDF rövidítés a Term Frequency-Inverse Document Frequency (magyarul: kifejezés gyakoriság-inverz dokumentum gyakoriság) kifejezésből származik és egy fontos módszer a dokumentumok közötti hasonlóság meghatározásában az információ visszakeresés és a természetes nyelvfeldolgozás területén. A kifejezés gyakoriság méri, hogy egy adott szó milyen gyakran fordul elő egy dokumentumban. A gyakoriság növekszik, ha a szó többször szerepel a dokumentumban. Az Inverz Dokumentum Gyakoriság pedig azt méri, hogy egy adott szó mennyire egyedi vagy jellemző a dokumentumhalmazra nézve. Az IDF magas értéket kap, ha a szó ritkán fordul elő más dokumentumokban.
- PLM:** A Pre-trained Language Model egy olyan mesterséges intelligencia modell, amelyet előzetesen nagy mennyiségű szöveges adattal tanítottak be, hogy megtanulja a nyelvi szerkezeteket, a szavak közötti kapcsolatokat és a

kontextusokat. Ezáltal a modell képes előrejelzéseket tenni szövegekkel kapcsolatban, illetve megérti a nyelvet és képes válaszolni különböző kérdésekre és megoldani feladatokat. Az ilyen modellek használatának előnye, hogy nem kell minden alkalommal nulláról kezdeni a betanítást, hanem a modell már rendelkezik előzetes tudással, amelyet könnyen fel lehet használni új feladatokra.

BEVEZETÉS

Az elmúlt években a mesterséges intelligencia felhasználása ugrásszerűen megnőtt, ezzel párhuzamosan a természetes nyelvfeldolgozás használata is rohamosan emelkedett. Ez a technológiai fejlődés lehetőséget teremt arra, hogy radikálisan átalakítsa az eddigi általános módszertanokat a tőzsdei elemzésekben. Az európai energia transzformáció 2020 óta jelentősen felgyorsult, ami köszönhető az elektromos autók nem várt mértékben történő elterjedésének és a szabályozói elköteleződésnek a fenntarthatóság iránt. Ezt a tendenciát tovább fokozta az orosz-ukrán konfliktus által kiváltott energiaválság is, ami még inkább előtérbe helyezte a megújuló energiaforrások iránti igényt. Eközben a befektetők körében egyre inkább elterjedt az a várakozás, hogy a pénzüket olyan vállalatokba fektessék, amelyek nem csupán eredményesek, hanem fenntarthatók is. Ez a folyamat soha nem látott nyomást gyakorol Európa legnagyobb olajvállalataira, hogy működésüket a fenntarthatóság irányába tereljék. Ugyan a vállalatok többsége már régóta közöl éves fenntarthatósági jelentést, viszont a negyedéves jelentésekben található fenntarthatósággal kapcsolatos információk nagyobb hatással lehetnek a részvényárfolyamokra, így dolgozatomban a negyedéves jelentéseket vizsgáltam.

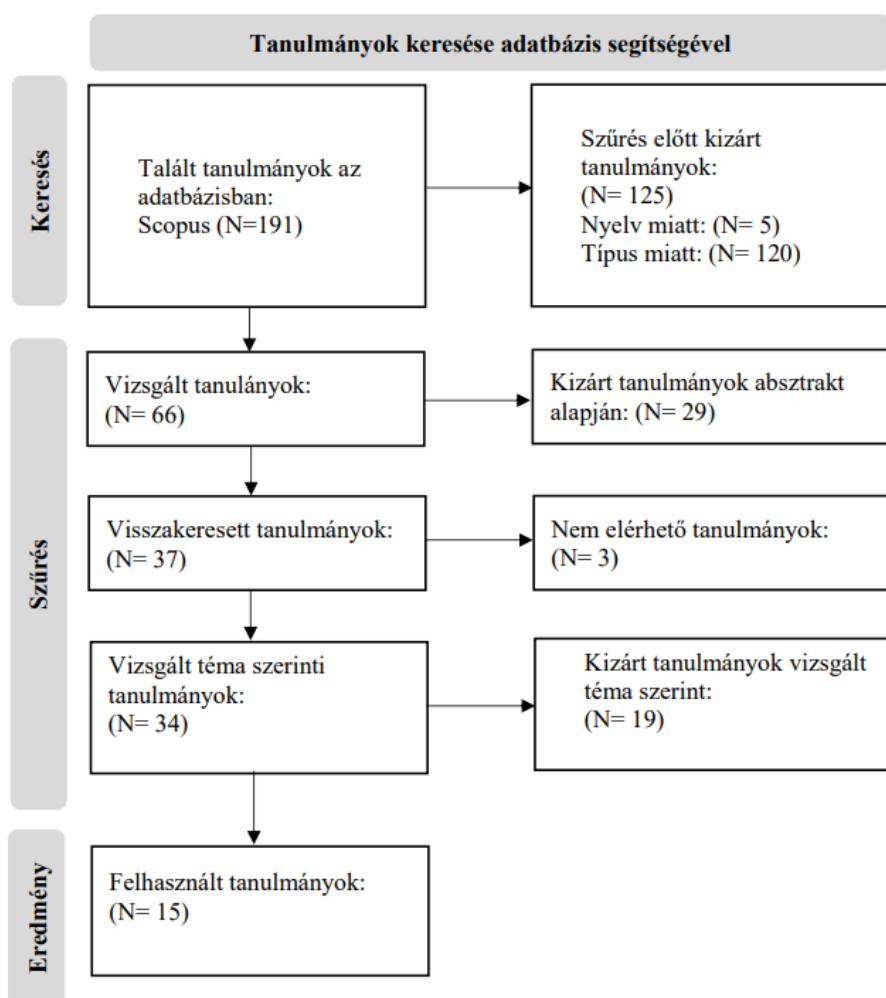
A dolgozat gyakorlati részét két részre bontottam, szekunder és primer kutatásra. Szekunder kutatásom keretében elemeztem Európa 10 vezető, főként integrált olajvállalatát, figyelembe véve szénhidrogén kitermelésüket, és kőolaj finomítási tevékenységüket. A vállalatok pénzügyi teljesítményét számadatok alapján is bemutatom, kiemelve tőkeerejüket, és pénzügyi helyzetüket. Emellett kifejtem azokat a tényezőket, amelyek magyarázatot adhatnak az egyes vállalatok pénzügyi mutatószámai közötti eltérésre.

Primer kutatásomban arra keresem a választ, hogy van-e kapcsolat az egyes olajvállalatok 2020. és 2023. közötti negyedéves jelentéseiben található fenntarthatósággal kapcsolatos szavak, és e cégek tőzsdei negyedéves árfolyam változásai között. A teljesebb kép érdekében, korrelációs vizsgálatot végeztem vállalatonként, az egyes negyedéves jelentésekben lévő fenntarthatósággal kapcsolatos szavak között. Mindemellett megnéztem, hogy van-e korreláció az elemzett vállalatok összes negyedéves jelentéseiben használt fenntarthatósággal kapcsolatos szavak között. Bár vannak tanulmányok, amelyek a negyedéves jelentések és az árfolyamok közötti kapcsolatot vizsgálják, viszonylag kevés kutatást találtam, melyek kifejezetten az olaj cégek részvényárfolyama, és az általuk riportált fenntarthatósággal kapcsolatos kijelentések közötti kapcsolatot keresnék. Dolgozatomban így erre a területre koncentrálok, és a kutatásomat egy lehetséges új elemzési eszközzel, a természetes nyelvfeldolgozással végzem el.

1. SPECIFIKUS INFORMÁCIÓ KÖZZÉTÉTELEK GAZDASÁGI HATÁSAINAK MÉRÉSE TERMÉSZETES NYELVFELDOLGOZÁS MÓDSZERÉVEL KIEGÉSZÍTVE

A szakirodalom feldolgozása során szisztematikus szakirodalmi gyűjtést végeztem, amelyhez a PRISMA munkafolyamat ábráját használtam fel. A gyűjtés során használt keresőkifejezés: natural AND language AND processing AND finance AND forecast OR stock OR market. A következő ábrán látható a szakirodalmi keresés és szűrés menete a megadott kulcsszavakkal. A szűrés végén 15, dolgozatom szempontjából releváns kutatást találtam, amelyek feldolgozásra kerültek. Mindemellett, ezen 15 tanulmány által hivatkozott más szerzők és magyar kutatók kutatási eredményeit is felhasználtam a minél teljesebb szakirodalmi áttekintés érdekében.

1. ábra
A szakirodalmi szűrés menete



Forrás: Moher és társai, (2009), saját szerkesztés

1.1. A TERMÉSZETES NYELVFELDOLGOZÁS SZAKIRODALMA

Az NLP-t, mint technológia ötletet nem a 21. században fogalmazták meg először, már az 1950-es években is folytak meghatározó kutatások és születtek elméletek arra vonatkozóan, hogy egy gép hogyan tudja értelmezni az emberi beszédet. Az 1960-as években King, Masterman, Ceccato és Yngve olyan fordítógépen dolgoztak, amely magát az NLP technológiát alapozta meg (Krutilla és Kővári, 2022). Az az elképzelés, hogy maga a nyelv, lexikonok halmaza és egyúttal szintaktikai rendszer (Bühler, 1934), már az NLP kialakulása előtt is felmerült. E megállapítással összhangban az NLP-vel való elemzés korai, népszerű megközelítései is olyan szemléletet képviseltek, amely vagy a kifejezőképességet (Sowa, 1987), vagy a nyelvi szabályokat (Chomsky, 1956) helyezte előtérbe. A mai napig a legtöbb NLP technika, amelyeket az NLFF-n alkalmaznak, továbbra is beilleszthető ebbe a két kategóriába, vagy ezek kombinációjába.

A legtöbb korai NLFF cikk a szöveges pénzügyi adatok számítógép által is könnyen feldolgozható tulajdonságoként való ábrázolására olyan szóhalmazt alkalmaztak, amelyek a szöveg szerkezetét a szavak halmazával és azok előfordulásának gyakoriságával reprezentálták. Ennek a technikának a hátránya, hogy nem veszi figyelembe a szavak sorrendjét. Ez a probléma bizonyos esetekben súlyos lehet. Például „a Samsung előnyhöz jut az Apple-lel szemben” és „az Apple előnyhöz jut a Samsung-gal szemben” hírek ellentétes hatást váltanak ki a tőzsdén, hiába használják ugyanazt a szóhalmazt. A másik probléma, amikor egy hír különböző szavakkal van megfogalmazva, mint például „A Brexit esést okozott a fontban” és „Az EU elhagyása felgyorsítja a font zuhanását”, ebben az esetben a szemantikai hasonlóság nem lesz figyelembe véve (Xing és társai, 2018).

Ezeket a problémákat jól kezeli egy szó és kontextusának figyelembevétele. A neurális hálózatok felhasználhatók arra, hogy tömören reprezentálják a szavakat (Bengio és társai, 2003). Ez a reprezentáció lehetővé teszi a szemantikai hasonlóságok számítását. A szó reprezentációk mellett a témamodellek (Blei, 2012) nagyléptékű gyűjtemények szemantikáját rögzítik. A dokumentum szintjén a szemantika több témára és azok megfelelő relevancia együtthatóira bontható. Ezek a technikák lehetővé teszik a nagy mennyiségű pénzügyi cikk teljes körű elemzését (Xing és társai, 2018).

Ebben a fejezetben bemutatom a legfontosabb természetes nyelvfeldolgozással (NLP) kapcsolatos szakirodalmakat. Ezen tanulmányok körébe tartoznak a különböző NLP módszerek és algoritmusok, a hangulatelemzés és a tőzsdei elemzések NLP-vel.

1.1.1. Természetes nyelvfeldolgozási módszerek és algoritmusok

Kamal és társai (2022) a következő megállapításokra jutottak: A BoW (szóhalmazok) és a TF-IDF (kifejezés gyakoriság – fordított dokumentum gyakoriság) módszereknek jelentős hiányosságai vannak, annak ellenére, hogy mind a BoW, mind a TF-IDF megszámolja a szavak előfordulásának számát a szövegben. A szó egyedisége és a szavak sorrendje nem kerülnek figyelembevételre, így képtelenek reprezentálni az írás jelentését és dimenzióit. A szóbeágyazási algoritmusokat különböző gépi tanulási módszerekkel elemezték, mint például az XGBoost, Naive Bayes, LSTM és CNN. Ezeket az eredményeket összehasonlították más kutatásokban alkalmazott módszerekkel. A legjobb megközelítések a BoW + CNN (80,86% pontosság), TF-IDF + CNN (80,53% pontosság), TF-IDF + LSTM (84% pontosság), BERT (90% pontosság) és a RoBERTa (88% pontosság) voltak. Fejlettebb megközelítés a BERT és a RoBERTa modellek alkalmazása, mivel teljesítményük felülmúlja a régebbi módszereket. Ezek a modellek képesek nagyobb szövegmennyiséget kezelni. Pontosságuk 90% illetve 88%, ami jobb, mint néhány gépi tanulási modell és más mélytanulási modellek, mint például az LSTM és a CNN.

Park és Cho (2023) kutatásukban bemutattak egy keretrendszert, amely egy újszerű vállalatnév maszkolási módszert alkalmaz a PLM-ek (nyelvi modellek) pénzügyi szakterület hatékonyságának növelésére. A széleskörű kísérleti eredmények azt mutatták, hogy a javasolt vállalatnév maszkolási módszer gazdagítja a modell ténybeli tudását, ezáltal javítva a részvényekkel kapcsolatos spam szűrési feladatok teljesítményét, különösen a 400 vagy kevesebb tréningadat méretű környezetekben. A javasolt megközelítésük, amely bizonyos típusú entitásokra vonatkozó információkat automatikusan kinyer egy mondatból, könnyen alkalmazható más területeken, amelyek háttértudást igényelnek egy entitás megértéséhez. Ezt a megközelítést ki lehetne terjeszteni szabadalmi dokumentumokra, amelyek új technológiákat írnak le. Egy szabadalmi tudással kibővített nyelvi modell gyorsabban és pontosabban értékelheti az új szabadalmi kérelmek újdonságát, ezzel elősegítve a szabadalmi vizsgálati eljárások automatizálását.

1.1.2. Hangulatelemzés természetes nyelvfeldolgozással

A pénzügyi szövegek könnyen elérhetővé váltak az interneten történő posztolások elterjedése és a piaci átláthatóság iránti egyre növekvő igény miatt. Ezek a trendek hozták létre a pénzügyi szövegek elemzését. A hangulatelemzés a szöveg szerzőjének egy tárggyal kapcsolatos álláspontját vizsgálja. Tipikusan a szerző attitűdjének, véleményének, értékelésének vagy a tárggyal kapcsolatos érzelmi megnyilvánulásának megtalálását célozza. A hangulatelemzés az

üzleti szféra talán legnépszerűbb NLP módszere, használják például a közösségi média hozzászólásaiban adott, a céggel vagy a cég bizonyos termékeivel, mozifilmekkel stb. kapcsolatos vélemények detektálására (Németh és társai, 2020). Klein és Prestbo (1974) felmérése bemutatja, hogy egy pesszimista pénzügyi híradás milyen hatással lehet a piacokra, és ez a tanulmány alátámasztja azt a javaslatot, hogy a híradások és a piacok kölcsönösen befolyásolják egymást (Chan és Chong, 2017).

Eddig két irányban zajlottak kutatások, melyek a tőzsdei piacok előrejelzését célozták meg hangulatelemzés felhasználásával. Az elsőben a kutatók kinyerték a hangulattípust (pozitív, negatív vagy semleges) a dokumentumokból, mint tulajdonságok (Wang és társai, 2019), a második irányban pedig a nyilvános hangulat idősorát számolták ki a hangulatértékek időszakonkénti összesítésével (Yang és társai, 2019). A piacon két kategória létezik a pénzügyi hangulatértékelés (FSA) terület-specifikus felhasználásában: aprólékos és általános. Az aprólékos hangulatértékelés a céggel kapcsolatos specifikus mondatrészekben belüli feladatokat foglalja magába, míg az általános hangulatértékelés gyakran dokumentumonként vagy mondatonként végez elemzést. A hangulatértékelés szempontjából két kategóriára oszthatók az alkalmazott megközelítések: lexikon-alapú és gépi tanulás-alapú módszerek. A lexikon-alapú módszerek a szöveg hangulattartalmát elemzik a hangulatszótár és a szópolaritás magas minősége alapján, míg a gépi tanulás-alapú megközelítések a felügyelt tanulásra és a szójellemzőkre támaszkodnak. A lexikon-alapú megközelítésekben a terület-specifikus vagy domain-adaptációs lexikon előre definiált manuális szabályokon alapul. Ilyen lexikonok például a Sentiwordnet (Baccianella és társai, 2010) és a Loughran–McDonald (Loughran és McDonald, 2011), amelyek domain-adaptációs és pénzügy-specifikus lexikonok a pénzügyi érzelemértékelésben. Azonban olyan tanulmányok, mint például Oliveira és társaié (2016), bemutattak egy módszert a pénzügyi domain-specifikus lexikon létrehozására gépi tanulás-alapú technikákkal. A gépi tanulás-alapú módszereknél a felügyelt módszereket vizsgálják a hangulatértékelésre, majd a dokumentum-, mondat- vagy aspektum-szintű hangulatérték alapján az in-domain osztályozót képezik az hangulatértékelési fázisban (Farimani és társai, 2022).

A mélytanuláson alapuló hangulatelemzési modellek tervezése és építése jelentős adathalmazokat igényel a képzéshez és teszteléshez. Bár számos nagy, nyilvánosan elérhető hangulat-címkézett adathalmaz létezik, ezek leginkább termékekhez és filmekhez kapcsolódnak. Sok hangulatelemzési modell (Yenter és Verma, 2017) használja ezeket az adathalmazokat és jó teljesítményt ér el azonos területeken. Azonban ezeknek a modelleknek

az alkalmazása különböző területeken kihívást jelent, mert mindegyik területnek saját érzelmi kifejezésszótára van (Mishev és társai, 2020).

A pénzügyi területet egyedi szókincs jellemzi, ami speciális, területspecifikus hangulatelemzést igényel. A pénzügyi piacokon megfigyelt árak minden elérhető információt tükröznek a kereskedett eszközökkel kapcsolatban (Malkiel, 2003), ezért az új információk lehetővé teszik a résztvevők számára, hogy jól tájékozott és időben meghozott döntéseket hozzanak. A hírekben és tweetekben kifejezett hangulatok befolyásolják a részvényárfolyamokat és a márká hírnevét, így ezeknek az érzelmeknek folyamatos mérése és követése egyre fontosabb tevékenységgé válik a befektetők számára. Bizonyos tanulmányok a hangulatelemzés alapján pénzügyi hírek felhasználásával jelzik előre a részvényárfolyamokat (Souma és társai, 2019), valutaváltást és globális pénzügyi piaci trendeket (Crone és Koeppel, 2014), valamint vállalati eredményeket (Mishev és társai, 2019).

Mivel a pénzügyi szektor saját szakzsargonra épül, nem elég általános hangulatelemzést alkalmazni a pénzügyek területén, mert sok szó eltér általános jelentésétől. Például a "felelősség" általában negatív szó, de a pénzügyi területen semleges jelentéssel bír. A "részvény" kifejezés általában pozitív értelmet hordoz, de a pénzügyi területen részvény pénzügyi eszközt vagy értékpapírt jelent, ami semleges jelentésű. Emellett a "bika" általában semleges, de a pénzügyekben szigorúan pozitív, míg a "medve" általában semleges, de a pénzügyekben negatív. Ezek a példák hangsúlyozzák a dedikált modellek fejlesztésének szükségességét, amelyek képesek kinyerni az érzelmeket a pénzügyi szövegekből (Mishev és társai, 2020).

1.1.3. Természetes nyelvfeldolgozás és gépi tanulás alkalmazása a tőzsdei elemzésekben

A tőkepiaci kutatás alapja a fundamentális és technikai elemzés (Isaac Kofi és társai, 2020). A fundamentális megközelítések magukban foglalják a részvénytőzsi előrejelzéseket numerikus adatok, például árfolyamok felhasználásával. A technikai megközelítések pedig a pénzügyi események időbeli dimenziójára összpontosítanak. Ezek a megközelítések historikus adatokra alkalmaznak trendmodellezési technikákat, hogy előre jelezzék az árfolyam alakulását. Korábbi kutatások pénzügyi híreket (Atkins és társai, 2018), blogokban és a közösségi médiában lévő befektetési hozzászólásokat (Swathi és társai, 2022) vettek figyelembe. Ezek a rendszerek természetes nyelvfeldolgozási (NLP) technikákat (Xing és társai, 2018) vagy gépi tanulási (ML) modelleket (Delic és társai, 2019) alkalmaztak, amelyek lehetnek felügyelték (Rustam és társai, 2020), automatikusan vagy kézi címkézéssel adathalmazokra támaszkodók, vagy nem felügyelték (Solorio-Fernandez és társai, 2020), figyelembe véve a bemeneti adatok

sajátosságait és leíró mintázatait. A legegyszerűbb megközelítés a tartalom vektoros reprezentációjának használata, ahol értéktelen elemek, mint például előjelek (García-Méndez és társai, 2020), eltávolításra kerülnek. Még komplexebb megközelítések, mint amilyen a De Arriba-Perez és társai (2020) által bemutatott kutatás, igyekeznek azonosítani a szintaktikai és szemantikai mintákat a pénzügyi hírek kulcsszávaiban lexikai, grammatikai és névazonosítási technikák segítségével. A tőkepiaci szűrés adatbányászattal (data mining) jelentős kutatási eredményeket hozott, amelyek hatékonyan támogatják a befektetési döntéshozatalt (García-Méndez és társai, 2023).

Alanyali és társai (2013) kimutatták a tőkepiaci események és a nyilvános adatok közötti kapcsolatot a pénzügyi hírekben. Atkins és társai (2018) a Latent Dirichlet Allocation gépi tanulási modelleket alkalmazták a tőkepiaci volatilitás előrejelzésére. Arra a következtetésre jutottak, hogy a hírek forrásaiból kinyert információk pontosabbak voltak a pénzügyi eszközök volatilitásának előrejelzésére, mint az árváltozások. Más szerzők mély tanulási algoritmusokat alkalmaztak pénzügyi hírekre (Vargas és társai, 2017). Konkrétan Day és Lee (2016) arra a következtetésre jutottak, hogy a források különleges típusának (hírek, blog hozzászólások stb.) jelentős hatása van a végleges döntésre, míg Vargas és társai (2017) bebizonyították, hogy a visszatérő neurális hálózatok jobbak az összetett időbeli jellegzetességek befogásában és modellezésében a tőkepiaci előrejelzésekhez. Annak ellenére, hogy számos kutatás foglalkozott a tőkepiaci ismeretek kinyerésével, az időbeli dimenzió elemzése figyelmet érdemel, mivel a legtöbb szakirodalomban (Rickett, 2016) az időbeliség kizárólag a hírek, posztok vagy hozzászólások időbélyegeiből származik. A korai kutatások (Schumann, 1987) különböző nyelvi perspektívákból vizsgálták az időbeliség kifejezését: morfológia, szemantika és pragmatika. Ezenkívül Forray és Woodilla (2005) idővel kapcsolatos szavakat szűrtek ki a folyóirat címeiből (García-Méndez és társai, 2022).

A pénzügyi piacok és a közzétett tartalom közötti kapcsolatot nemcsak az egyedi részvények esetén találják meg, hanem a részvényindexek esetén is. Például az S&P 500 index (Zubair és Cios, 2015) pozitív korrelációt mutat a kifejezetten konstruált hangulatmérővel. Előrejelző beállításban a médiacikkek elősegítik a kutatásokat, amelyek megjósolják a Dow Jones index aznapi hozamát (Siering, 2013). Hasonlóképpen a hírhangulat lendülete is képes megjósolni a CDAX mozgásának irányát (Hagenau és társai, 2013).

Számos korábbi tanulmány vizsgálta a részvényindexek előrejelzési lehetőségét csupán a historikus értékeken alapulva, és ezért csak néhány példát említek a következőkben. Előrejelzési kísérleteket végeztek különböző indexekre, mint például az S&P 500-ra (Rounaghi és Zadeh, 2016), az NYSE-re (Pai és Lin, 2005), a Dow Jones Ipari Átlagra (Wang és társai,

2012), a NIKKEI 225-re (Huang és társai, 2005), valamint a feltörekvő piacokra (Oztekin és társai, 2016). Ezek a munkák gyakran idősor elemzési módszereket alkalmaztak, mint például autoregresszív vagy mozgóátlagos folyamatok, időről időre a volatilitásmodellezési megközelítésekkel kombinálva (Herwartz, 2017). Azok a munkák, amelyek közel állnak a gépi tanuláshoz, kísérleteznek például a támogató vektor algoritmusokkal, neurális hálózatokkal és a neurális hálózatokat és autoregressziót kombináló hibrid modellekkel (Feuerriegel és Gordon, 2018), ezért mind autoregresszív, mind nem-lineáris gépi tanulási modelleket használtak historikus értékekkel kiindulási alapként (Feuerriegel és Gordon, 2018).

1.2. A TERMÉSZETES NYELVFELDOLGOZÁS LEHETSÉGES ALKALMAZÁSI TERÜLETEI

A természetes nyelvfeldolgozás rendkívül sokoldalúan alkalmazható különféle területeken, mint például a pénzügyekben, ahol a különböző szövegek elemzésével előrejelzéseket készíthetünk a segítségével. Ebben a fejezetben bemutatom a természetes nyelvfeldolgozás dolgozatom szempontjából releváns alkalmazási területeit.

1.2.1. Részvénytőke volatilitásának előrejelzése

Atkins és társai (2018) empirikus eredményeket mutattak be, amelyek szerint az információk, amelyeket szöveges hírforrásokból nyertek ki, felhasználhatók a piaci volatilitás irányának előrejelzésére. Rávilágítottak, hogy a volatilitás változásai jobban előre jelezhetők, mint egy eszköz vagy eszközindex záróárainak változásai. Az eredményeik szerint a hírekben található információk, amelyek a piacot az érzelmezérelt viselkedésen keresztül befolyásolják, alapvetően hatnak a pénzügyi rendszer másodrendű statisztikáira.

Az eredmény, hogy a másodrendű statisztikák jobban kapcsolódnak a hírekhez, valóban kihasználható a kereskedés során, a derivatívák árazása révén. Kutatásuk a hírek által származtatott információkra és az alapul szolgáló részvények számára becsült volatilitások közötti kapcsolatra összpontosított, amelyeket opciós árazási képletek inverziójával számítottak ki (pl. Black-Scholes).

Fontos megjegyezni, hogy az általuk jelentett előrejelzési eredmények kizárólag a hírforrásból származó információkra alapulnak és más bemenetet nem használtak. Ezért a pontos eredményeket nem lehet közvetlenül összehasonlítani azokkal az eredményekkel, amelyek az idősorokra alapozott modellek esetében készülnek (előrejelzések a múltbeli értékekből). Kiemelendő, hogy a hírekből származó információk egyedül is tartalmaznak egy

olyan jelet, amely elég erős ahhoz, hogy megjósolja a piaci volatilitás irányát. Nyilvánvaló, hogy egy modell, amely akár csak a korábbi volatilitási értékeket is magában foglalja, valószínűleg magasabb pontossági szinteket mutatna. (Atkins és társai, 2018)

A pénzügyi piacok adatai kihívást jelentő jelzésfeldolgozási feladatokat kínálnak és nagy érdeklődést váltottak ki a kvantitatív kutatókból és kereskedőkből. Az általuk generált pénzügyi rendszer alapvető bonyolultsága hatalmas, így a rendszerből származó adatok non-lineáris és nem-stacionárius tulajdonságokat mutathatnak. Az elektronikus adatok könnyű elérhetőségével a közelmúltban jelentős növekedés volt tapasztalható az elemzések terén, és ezek az elemzések automatizált kereskedési rendszerekbe történő átültetését is magukkal hozták. Becslések szerint az algoritmikus kereskedések 2012-ben a piaci forgalom 85%-át tették ki, szemben a 2003-as 15%-kal. A derivatívák piaca is jelentősen nőtt, az Egyesült Államokban jegyzett részvény- és indexopciók átlagosan napi 16,5 millió szerződést kötöttek az első öt hónapban 2017-ben, és a globális tőzsdén jegyzett opciós piac 2016 decemberében több mint 41 ezer milliárd dollárra volt értékelve. (Atkins és társai, 2018)

A portfóliókat optimalizálni lehet a pénzügyi eszközök (például részvények) várható hozamainak és ármozgások bizonytalanságainak alapján, valamint a derivatívák és más komplex szerződések árazása valamilyen jósolt jövőbeli eszköz alapján (például a Black-Scholes-modell) és általános idősor elemzéssel a sokféle egyváltozós és többváltozós analitikai technika felhasználásával, többek között a statisztika, a kvantitatív pénzügy, a jelrendszer feldolgozás és a gépi tanulás területén. Ezek a kvantitatív modellezés eszközei, amelyeknek az alapvető formái a technikai kereskedés gyakorlatában is megtalálhatóak, kiegészítik azt, amit alapvető elemzésnek nevezünk, amely során a vagyonelemekről szóló kvalitatív információkat makrogazdasági információkkal kombinálva használunk fel az eszköz értékére vonatkozó előrejelzések készítéséhez. (Atkins és társai, 2018)

A szerzők által bemutatott újítás az, hogy az eszközárak és a volatilitás változásainak előrejelzését kizárólag a hírekből származó információk alapján végezték, és nem integrálták az idősoros modellt. Ez az eredmény a volatilitás előrejelzését eredményezi, amely egy kis, ugyanakkor statisztikailag jelentős jelet hoz létre. Mindemellett kutatásukat sokkal nagyobb adathalmazon végezték el, mint amit az eddigi szakirodalmakban szoktak. (Atkins és társai, 2018)

1.2.2. Jegybankelnöki beszédek vizsgálata

Park és társai (2023) azt vizsgálták, hogy lehetséges-e pénzügyi válságnak megfelelő kockázati tényezők korai felismerése a Federal Reserve Bank elnökeinek beszédeiben történő hot topic-

ok keresésével. Egy szövegbányászati keretrendszert javasoltak olyan potenciális piaci kockázati tényezők azonosítására, amelyeket a Federal Reserve Bank elnökeinek beszédeiben történő hot topic-ok keresésével lehet kinyerni. A keretrendszerben egy könnyű, felügyelet nélküli kulcsszó-pontszámítási módszert alkalmaztak. Ez lehetővé tette a kulcsfontosságú kockázati tényezők hatékony felismerését egy pénzügyi válság előtt. A javasolt keretrendszert 491 jegybankelnöki beszéd értékelésére használták fel 1997 januárja és 2019 szeptembere között.

A keretrendszerből kinyert kulcsszavakat, mint például "Kelet-Ázsia", "információs technológia" és "adósság fedezeti értékpapír", amelyek leírták az ázsiai pénzügyi válságot, az internetes buborékot és a globális pénzügyi válságot a válságok előtti beszédekből. Emellett a kvantitatív értékelés igazolta a válságok korai felismerésének lehetőségét a jegybankárok beszédeinek használatával. Több, mint 400%-os növekedés mutatkozott meg a releváns szavak gyakoriságában a jegybankárok beszédeiben a hírekkel szemben. Ez arra utal, hogy a jegybankárok beszédei több hasznos információt tartalmaznak a pénzügyi válságok előrejelzéséhez, mint a hírek. (Park és társai, 2023)

A javasolt megközelítéshez specifikus területi stopszavak szükségesek, amelyeket a kutatóknak gondosan kell kiválasztaniuk a jelentéktelen kifejezések kiszűrése érdekében a hot topic-ok kereséséhez. Az előre meghatározott stopszavak készítése azonban munkaigényes és szubjektív folyamat. A kutatók beavatkozásának csökkentése érdekében fontolóra vehető a területspecifikus stopszavak automatikus azonosítása a jövőben. Azt is érdemes lenne tovább vizsgálni, hogy milyen mértékben kell tükrözni a múltat a pontszámítási folyamatban, mint jövőbeli tanulmány. A javasolt "megjelenési" pontszám a kulcsszavak időbeli fontosságát veszi figyelembe a korábbi négy negyedévben használt szógyakorisággal; azonban szükség lehet egy rövidebb vagy hosszabb időszak figyelembevételére is. Bár a javasolt keretrendszer teljesítménye ígéretes, van még fejlődési lehetőség. Például az elemzés célterületének hatókörét ki lehet terjeszteni más típusú jegybanki kommunikációkra, például a Federal Open Market Committee nyilatkozataira vagy jegyzőkönyveire. Ez lehetővé tenné, hogy összehasonlítsák az egyes dokumentumtípusok hasznosságát a pénzügyi válságok kockázati tényezőinek előzetes azonosításában. Mindemellett lehetővé válik az elemzési terület kiterjesztése az Európai Központi Bank vagy más országok jegybankjainak beszédeinek elemzésével. (Park és társai, 2023)

1.2.3. A közvélemény és politikai helyzet hatása a tőzsdére

Khan és társai (2020) azt vizsgálták, hogy a közvélemény és a politikai helyzet hatása a gépi tanulási algoritmusokkal mennyire pontos a különböző tőzsdei piacokra vonatkozóan. A szerzők azt is kimutatták, hogy összefüggések vannak különböző részvények között. A véleményattribútum hozzáadásával a szerzők megállapították, hogy a nyilvános vélemény hatása egy adott vállalatra minimális. Ennek két lehetséges oka van: vagy a véleménynek nincs hatása, vagy a jelenleg létező véleményelemzési technikák nem nyújtanak elég információt a kapcsolat megértéséhez a vélemény és egy vállalat részvénye között. A szerzők megállapították, hogy a hatás 0-tól 3%-ig terjedt. A véleményattribútum továbbá körülbelül 2%-kal javította a gépi tanulás algoritmusainak pontosságát. Emellett a tőzsdei trendek a kereskedés napját követő hetedik napon kerültek előrejelzésre a legnagyobb pontossággal.

Nagyon fontos következtetés ebből a kutatásból a politikai események hatása egy ország tőzsdéjére, ez a hatás 10% és 20% közötti. A szerzők 98 politikai eseményen tesztelték algoritmusukat, és az előrejelzési pontosság javulása jelentős volt. A legmagasabb előrejelzési pontosság azt mutatta, hogy a politikai helyzet hatása a leginkább hatékony az esemény bekövetkezésétől számított 5. napon. (Khan és társai, 2020)

Az alkalmazott kutatások szempontjából a JB és a SNLP megközelítések hasonló előrejelzési eredményeket mutattak. Nem volt olyan algoritmus, amely minden vállalat részvényeinek árát meg tudta volna jósolni a legmagasabb pontossággal a véleményelemzés felhasználásával. Ugyanakkor a politikai helyzet elemzésében a szerzők azt találták, hogy a politikai helyzetek bevonása megközelítőleg 20%-kal javította az MLP és a DT előrejelzési pontosságát. (Khan és társai, 2020)

Khan és társai (2020) kutatása a Schumaker és Chen módszer pontosságát is javította (Schumaker és Chen, 2009), amely eredetileg 57,1% volt különböző részvénypiaci adatokat felhasználva. A javasolt algoritmus akár 68%-os pontossággal jósolta meg az eredményeket és mindez hét munkanappal később történt. A politikai összefüggés szempontjából a tanulmány kihasználta a Taimur és Khan (2015) által talált korrelációt és egy gyakorlati gépi tanulási modellt fejlesztett ki, amely 75,38%-os pontosságot eredményezett, amit korábban soha nem értek el. Végül a szerzők bebizonyították, hogy pozitív összefüggések vannak azok között a vállalatok közötti tendenciák között, amelyek egy adott ágazathoz tartoznak. Ezen felül összehasonlító tanulmányt végeztek a jelenlegi modell és a korábbi modellek között. A összehasonlítás során megállapították, hogy a javasolt modell jobban teljesít, mint a meglévő modellek. (Khan és társai, 2020)

A kifejlesztett modellt ajánlani lehet befektetőknek és pénzügyi elemzőknek. A prognózis modell segíthet nekik előre látni a részvények jövőbeli viselkedését, és eldönteni, melyik részvényeket vásárolják meg és melyeket adják el. A modellt döntéstámogató rendszerekben is lehet használni. Például klinikai döntéstámogató rendszerben a modellt felhasználhatják a betegek orvosi feljegyzéseinek kritikus és nem kritikus betegségi osztályokba való besorolásához. A modellt felhasználhatják a vállalkozások és a menedzsment területén is a negatív és pozitív tendenciák azonosításához és a vállalkozási erőforrások jobb elosztásához. (Khan és társai, 2020)

1.2.4. Közösségi média üzenetek és a tőzsdei árfolyamok kapcsolata

Liu és társai (2023) a tőzsdei befektetők közösségi média üzeneteinek felhasználásával és NLP (természetes nyelvfeldolgozás) technikákkal mérték a befektetői hangulatot a közösségi hálózatokon és vizsgálták az alapvető mintázatokat a tőzsdei árak és a befektetői hangulat közötti összefüggésben. Megállapították, hogy általánosságban véve a befektetői hangulatnak gyenge, de rendkívül jelentős pozitív korrelációja van a tőzsdei árakkal. Azonban konkrét időszakokban (például egy vagy két hónapban) a befektetői hangulat és a tőzsdei árak közötti összefüggés lehet pozitív, negatív vagy semleges a pozitív és negatív együttműködési jelek kombinált hatása alatt.

Egy szegmentált befektetői hangulatomérés felhasználásával megállapították, hogy a konkrét részvényindexeknek nincs magasabb összefüggése a valós idejű közösségi média hangulattal. Ez azt sugallja, hogy a magas gyakoriságú közösségi média hangulatomérések lehet, hogy nem hasznosak a részvényárfolyamok előrejelzéséhez, ami összhangban van Behrendt és Schmidt (2020) eredményeivel. Emellett megállapították, hogy a nyitás előtti befektetői hangulatnak nagyobb hangulati összefüggése van a részvényárfolyamokkal. Ez azt mutatja, hogy a közösségi média zajosabb a nyitás után és, hogy a nyitás előtti befektetői hangulat relevánsabb a részvényárfolyamok előrejelzéséhez. Emellett arra is utal, hogy a befektetői hangulat a tőzsdén előretételező és ex ante információt tartalmaz, amely jelzi, hogy a jövőbeni árak statisztikailag szignifikánsak. (Liu és társai, 2023)

A természetes nyelvfeldolgozó technológia segítségével Liu és társai (2022) osztályozták a SSE tőzsdei sávban található 2,66 millió üzenet hangulatát 2019. január 1-jétől 2021. december 31-ig. Ez volt az alapja annak, hogy létre tudják hozni a befektetői bizalomindexet (ICI), mint egy hangulat helyettesítőt és kombinálják a befektetői hangulatot és a tőzsdei idézeteket annak elemzéséhez, hogy hogyan ingadozik a befektetői hangulat és mi az alapvető tényezők hatása a hangulat ingadozására. Az eredmények azt mutatták, hogy a tőzsdén a befektetői hangulatnak

van egy 7 napos ciklusa. Az 5 munkanapos kereskedési napokon a befektetők "racionális emberek" jellegzetességeket mutatnak a nyereség realizálása és a kockázatkerülési preferenciák révén. A 2 napos ünnepnapok alatt a befektetők "társadalmi emberek" jellegzetességeket mutatnak, a befektetői hangulat jelentősen magasabb és az ingadozások szűkebbek, ennek a növekedésnek a mértéke nem függ a tőzsdei volatilitástól, ami az ünnep hatást mutatja. (Liu és társai, 2022)

Számos kutató igazolta a befektetői hangulat és az olyan gazdasági jelenségek közötti kapcsolatot, mint a "tőzsdei hozamok", "részvényárak" és "tőzsdei likviditás" (Al-Hajieh és társai, 2011; Teng és Liu, 2013), különösen a "hétfő hatás" a tőzsdén (Gibbons és Hess, 1981; Lakonishok és Levi, 1982). A Liu és társainak (2022) eredményei bizonyos mértékig megerősítik ezeket a következtetéseket. Megállapították, hogy egy tipikus kereskedési ciklusban a kereskedési napok és ünnepnapok ciklikus mozgása $5 + 2$ ingadozást eredményez a befektetői hangulatban. A tipikus kereskedési ciklus mintázat egy potenciális magyarázat lehet sok befektetői hangulathoz kapcsolódó gazdasági jelenség esetén.

Az öt kereskedési nap és két ünnepnap kereskedési ciklusa azonban a befektetői hangulat ingadozásának elsődleges oka. A bonyolult külső okok eltávolíthatják a befektetői hangulatot a grund volatilitási ciklusból, de amikor a külső erők eltűnnek, a befektetői hangulat visszatér az alap volatilitáshoz. A kereskedési napok és ünnepnapok ciklusa napról napra változtatja a befektetői hangulatot, ami viszont meghatározza a hét napjának hatását a tőzsdei hozamokra. Ez egy átfogó magyarázat a kutatók részéről az ünnepnapok és a befektetői hangulat DOW hatásának értelmezésére. Eredményeik azt sugallják, hogy az ünnepnapoknak fontos hatása van a befektetői hangulatra és, hogy a befektetői hangulat az ünnepek idején jelentősen magasabb, ami hatással lehet a befektetői hangulat megértésére és a tőzsdei rendszerek tervezésére. (Liu és társai, 2022)

1.2.5. A természetes nyelvfeldolgozás jövője

A természetes nyelvfeldolgozási (NLP) algoritmusokat, mint az egyik főáramú MI technikát, az elmúlt évtizedekben széles körben felhasználták az üzleti és pénzügyi területeken. Ezeknek a tanulmányoknak a többsége szöveges információkat használ, összefüggéseket keresve a magas gyakoriságú szövegek és a feltörekvő trendek között. Például Herzenstein és társai (2011) megmutatták, hogy a narratívák jelentősen befolyásolják a hitelezési döntéseket a különböző szöveges kontextuális információk terjesztésével. Agarwal és társai (2016) azt találták, hogy a nyelvi hang árnyalata a hitelminősítési jelentésekben szorosan kapcsolódik az anomális hozamokhoz. Hassan és társai (2019) kiterjesztett számítási módszert használtak a

nyereségkonferenciák elemzéséhez, hogy megmutassák, hogy a vállalatok politikai kockázatai szorosan kapcsolódnak részvényhozamaikhoz és beruházásaikhoz. A legújabb NLP algoritmusok már alkalmazást nyertek az üzleti és pénzügyi tanulmányokban (Korangi és társai, 2023; Stevenson és társai, 2021). A generatív MI, ellentétben a hagyományos NLP modellekkel, könnyedén létrehoz interakciókat, tervezéseket és akár hamis szövegeket is. Például Brynjolfsson és társai (2023) megmutatták, hogy a generatív MI nem igényel bemeneti utasításokat, mert felismeri a mintákat a tanító adathalmazból. Korábbi kutatások rámutattak arra, hogy a generatív MI alkalmazása a vállalkozási és pénzügyi területen még gyerekcipőben jár. Dowling és Lucey (2023) a leggyakrabban használt generatív MI eszköz, a ChatGPT felhasználásával új ötleteket generált és az adatokat azonosította a pénzügyi kutatásban. A kutatási eredmények arra utalnak, hogy a kutatói szakértelem és a bemeneti adatok szintje jelentős hatással lehet a ChatGPT kimeneti eredményeire. Rámutatnak továbbá ennek a technikának a relativitására az irodalom áttekintésének létrehozásában. A politikával kapcsolatos kutatás szempontjából Hansen és Kazinnik (2023) megvizsgálták, hogy a GPT és más fejlett NLP modellek mennyire képesek megfejteni a Fedspeak-et, azaz az amerikai jegybank monetáris politikai döntéseit részletező állításokat. Az emberi osztályozási benchmark alapján eredményeik arra utaltak, hogy a GPT modellek teljesítenek a legjobban. Ezzel párhuzamosan Yang és Menczer (2023) azt vizsgálták, hogy a ChatGPT alkalmazható-e a hírforrások hitelességének meghatározására. Eredményeik szerint a ChatGPT által generált értékelések szorosan korrelálnak az emberi szakértők véleményeivel. (Chen és társai, 2023)

2. TERMÉSZETES NYELVFELDOLGOZÁS MÓDSZERÉVEL PRIMER ÉS SZEKUNDER KUTATÁS EREDMÉNYEINEK BEMUTATÁSA

2.1. VÁLLALATOK BEMUTATÁSA ÉS SZEKUNDER KUTATÁS

Kutatásomhoz azért választottam az európai olajszektort, mert Európában van a legnagyobb nyomás az olaj cégeken, hogy működésüket fenntarthatóvá tegyék. Jelenleg ez a szektor szembeül a legnagyobb kihívással, így a befektetők valószínűleg nagy figyelemmel követik ezen vállalatok jövőre vonatkozó nyilatkozatait, amelyek a fenntartható működéssel

kapcsolatosak. Az elemzéshez Európa 10 legnagyobb tőzsdén jegyzett integrált olajvállalatát használtam fel.

Ebben a fejezetben bemutatom azt a 10 európai olajipari vállalatot, amelyek negyedéves jelentéseit természetes nyelvfeldolgozás segítségével fogom elemezni, kitérek többek között a legfontosabb pénzügyi adataikra és mutatóikra is, ehhez a vállalatok 2020. és 2023. közötti negyedéves jelentéseit használtam fel.

2.1.1. Vállalatok bemutatása

A következő táblázatban az elemzett vállalatok nemzetisége, az általuk kitermelt és finomított szénhidrogén 2023-as mennyisége látható.

1. táblázat
Elemzett vállalatok bemutatása

Megnevezés	Nemzetiség	Szénhidrogén kitermelés 2023 (ezer hordó/nap)	Kőolaj finomítás 2023 (ezer hordó/nap)
BP	Egyesült Királyság	2 312	1 411
Eni	Olaszország	1 655	536
Galp Energia	Portugália	122	216
MOL	Magyarország	90	235
Neste	Finnország	0	218
OMV	Ausztria	364	319
PKN Orlen	Lengyelország	190	860
Repsol	Spanyolország	599	823
Shell	Egyesült Királyság	2 791	1 349
TotalEnergies	Franciaország	2 483	1 436
Összesen	-	10 606	7 403

Forrás: negyedéves jelentések, saját szerkesztés

Az összes elemzett vállalat egy integrált olajipari cég a Neste kivételével. Ez azt jelenti, hogy működésük kiterjed az upstream, midstream és downstream tevékenységekre. Az upstream szegmens a kőolaj- és gázkitermeléssel foglalkozik, ideértve az olaj- és gázmezők felfedezését, fejlesztését is. A midstream szegmens általában az olaj- és gázvezetékek és terminálok üzemeltetésével foglalkozik. A downstream szegmensbe tartozik a vállalatok olajfinomítói és petrokémiai tevékenysége, valamint az olajtermékek (üzemanyagok, kenőanyagok, vegyi anyagok) forgalmazása és értékesítése. Ide tartozik továbbá az üzemanyag-kiskereskedelem, a töltőállomások üzemeltetése és a fogyasztók felé történő olajipari termékek értékesítése. Az integrált olajipari vállalatok e széles körű tevékenységi területeken való részvételük révén igyekeznek maximalizálni a hatékonyságot és optimalizálni a hozamot. A Neste-nek, ahogy a

táblázatból is látszik nincs upstream tevékenysége, mivel nem rendelkezik olaj vagy gázkitermeléssel, ezáltal nem tekinthető klasszikus integrált olajvállalatnak. A Neste a világ vezető gyártója a fenntartható repülőgép üzemanyagoknak, a megújuló dízelnek és a megújuló vegyipari alapanyag megoldásoknak. Mindemellett jelentős downstream tevékenységgel is rendelkezik, ami magába foglalja a kőolaj finomítást és az üzemanyag kiskereskedelmet.

Az elemzett 10 vállalat 2023-ben átlagosan naponta 10 606 ezer hordóval egyenértékű szénhidrogént termelt ki, ami a globális szénhidrogén termelés 6,13%-val egyenlő. A legnagyobb szénhidrogén termelők az elemzett cégek közül a BP, az Eni, a Shell és a TotalEnergies voltak. A 10 vállalat összesen átlagosan 7 403 ezer hordó kőolajat finomított naponta, ez az európai napi 12 379 ezer hordó összes finomított mennyiség 59,8%-a. A legjelentősebb kőolaj finomítók 2023-ben a BP, a Shell és a TotalEnergies voltak az elemzett vállalatok közül.

2.1.2. Elemzett vállalatok legfontosabb pénzügyi adatai

A következő táblázatban az elemzett vállalatok legfontosabb pénzügyi adatai találhatóak.

2. táblázat
Elemzett vállalatok 2023-as pénzügyi adatai milliárd dollárban

Megnevezés	Piaci kapitalizáció	Árbevétel	Profit	Eszközök összesen
BP	102,3	213,0	15,2	280,29
Eni	55,2	102,5	5,1	156,1
Galp Energia	12,0	22,3	1,3	18,1
MOL	6,4	26,0	1,5	22,3
Neste	27,3	25,1	1,6	17,5
OMV	14,3	43,2	1,6	55,4
PKN Orlen	19,3	90,0	6,7	70,5
Repsol	18,2	67,3	3,5	69,6
Shell	214,7	323,2	19,4	406,3
TotalEnergies	163,5	218,9	21,4	283,7
Összesen	633,2	1131,5	77,3	1379,8

Forrás: (companiesmarketcap.com 2024), negyedéves jelentések, saját szerkesztés

Jól látható, hogy mind piaci kapitalizációban, mind árbevételben és mind eszközértékben a legnagyobb vállalat egyértelműen a brit-holland Shell. Egyedül profitban előzi meg a francia TotalEnergies. 2023-ban az olajárak sokkal kiegyensúlyozottabbak voltak, mint 2022-ben. Míg 2022-ben 63-139 dollár között ingadoztak az árfolyamok, addig 2023-ban 65-97 dollár között. Az alacsonyabb olaj- és gázáraknak és finomítói árreseknek köszönhetően 2023-ban 24%-kal

csökkent az elemzett vállalatok profitja 2022-höz képest, amikor is rekord profitot értek el ezek a társaságok. A táblázatból az is kivehető, hogy e 10 vállalat értéke több, mint 600 milliárd dollár, összbevételük pedig meghaladja az 1100 milliárd dollárt. Összehasonlításképpen ez több, mint 6-szor nagyobb, mint Magyarország 2023-as GDP-je. Az 1. táblázat kitermelési, finomítási adataiból és a 2. táblázat pénzügyi adataiból is látszik, hogy az elemzett 10 olajvállalat lefedi az európai olajszektor nagy részét, így a kapott eredmények akár kiterjeszthetők más európai olaj cégekre is.

3. táblázat
Elemzett vállalatok 2023-as pénzügyi mutatószámai

Megnevezés	P/S	P/E	P/B	EV/EBITDA
BP	0,48	6,73	1,4	2,4
Eni	0,54	10,82	0,9	2,7
Galp Energia	0,54	9,23	2,6	4,4
MOL	0,25	4,27	0,5	2,1
Neste	1,09	17,06	2,3	8,7
OMV	0,33	8,94	0,7	1,8
PKN Orlen	0,21	2,88	0,5	1,6
Repsol	0,27	5,20	0,7	2,8
Shell	0,66	11,07	1,1	3,5
TotalEnergies	0,75	7,64	1,3	3,3
Összesen	0,51	8,38	1,2	3,3

Forrás: (finbox.com 2024), saját szerkesztés

A pénzügyi mutatókat vizsgálva szembetűnő, hogy a legnagyobb értékeltséggel a finn Neste rendelkezik az adózott eredményéhez képest. Ez annak köszönhető, hogy a Neste már 2015-ben elkezdte átalakítani üzleti modelljét, ennek első lépése volt a vállalat átnevezése Neste Oil-ről Neste-re, ezzel hangsúlyozva, hogy a vállalat ettől kezdve a fenntartható működésre kíván koncentrálni. A Neste élen jár az újrahasznosított anyagok és megújuló energiaforrások felhasználásában, továbbá a bioüzemanyagok és más fenntartható megoldások fejlesztésében is. 2023-ban a vállalat működési eredményének a 34%-a már megújuló termékekből származott. Mindemellett észrevehető, hogy más vállalatok, ahol ugyanúgy elindult a működés átalakítása és már jelentős eredmény termelő képessége van a fenntartható tevékenységeknek, magasabb szorzókon mozognak, mint azon szektortársaik, akiknek a profitja szinte teljes egészében a fosszilis energiahordozókból származik. Ilyen például a Shell és a TotalEnergies is, akiknek a gáz szegmensük (köszönhetően az LNG kereslet növekedésének is) már jelentős eredmény hozzájárulást ért el 2023-ban. A 3. táblázatból összességében megállapítható, hogy

a piac magasabbra értékeli azon olaj cégeket, akiknek van olyan fenntartható tevékenysége, amely már marginálisan képes hozzájárulni a vállalat eredményéhez.

2.2. PRIMER KUTATÁS

Ebben a fejezetben bemutatom a primer kutatásom vizsgálati szempontjait, eredményeit és ezek elemzéseit. Az áttekintés után következtetéseket vonok le és javaslatokat teszek a vizsgált kérdést illetően.

A primer kutatásomat négy részre bontva mutatom be. Az első rész a negyedéves jelentések elemzése, amelyben a Provalis Wordstat 2024 szoftver használatával vizsgálom a 10 vállalat 2020. és 2023. közötti negyedéves jelentéseit. A második részben bemutatom a kapott eredmények leíró statisztikáját, majd a 3. részben elvégzem a jelentésekben talált szavak közötti korreláció vizsgálatát. A negyedik részben pedig választ adok a dolgozat fő kutatási kérdésére, miszerint van-e kapcsolat az európai olajvállalatok negyedéves jelentéseiben található fenntarthatósággal kapcsolatos szavak száma és a részvényárfolyamok emelkedése, illetve csökkenése között.

2.2.1. Negyedéves jelentések elemzése

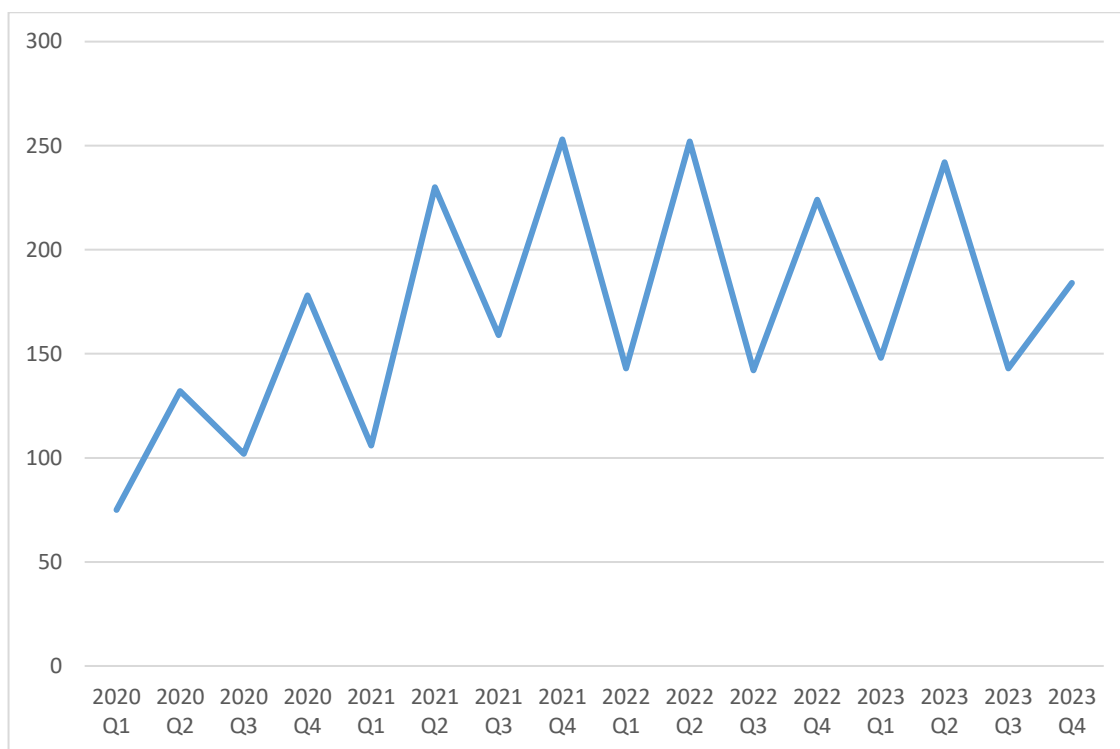
Az összesen mintegy 160 negyedéves jelentés elemzését a Wordstat szoftverrel végeztem el. Mindemellett felhasználtam Pencle és Mălăescu (2016) tanulmányát, amiben készítettek egy szótárat, amiben megtalálhatóak a legfontosabb fenntarthatósággal kapcsolatos szavak.¹ E szavak egyben indikátorok is arra, hogy meghatározzuk, hogy az adott negyedéves jelentésben mennyiszer beszélnek a fenntarthatóságról. Pencle és Mălăescu (2016) három listára bontották szótárjukat, az elsőbe a gazdasággal, a másodikba a környezettel, a harmadikba pedig a társadalommal kapcsolatos szavak kerültek.

Elemzésemhez a második, környezettel kapcsolatos listát használtam fel. Ennek oka, hogy az olajvállalatok esetében a legfontosabb szempont a környezeti, ami az általánosságban szennyező működésükből adódik, így a befektetők figyelme leginkább erre a területre koncentrál. Ahogy a szekunder kutatásomból is kiderült, azok a vállalatok, amelyek rendelkeznek a méretükhöz képest jelentős fenntartható tevékenységgel, azokat magasabbra értékeli a piac, mint azokat, amelyek nem.

¹ Például: emissions, green marketing, recycling, zero waste stb.

Az elemzés során a listában lévő 176 szóból 42 szó volt megtalálható a 160 negyedéves jelentés valamelyikébe. A következő táblázatban látható, hogy negyedévekre lebontva hány darab fenntarthatósággal kapcsolatos szó volt összesen a jelentésekben.

2. ábra
Fenntarthatósággal kapcsolatos szavak száma a negyedéves jelentésekben



Forrás: saját szerkesztés

Megfigyelhető, hogy 2020 óta a jelentésekben megnövekedett a fenntarthatósággal kapcsolatos szavak száma. Mindemellett az is kiderül, hogy a második és negyedik negyedéves jelentésekben általánosságban jelentősen magasabb a szavak száma. Ez abból adódik, hogy rendszerint a második és negyedik negyedéves jelentések részletesebbek és terjedelmük is hosszabb, mint az első és harmadik negyedévéké, ennek köszönhetően a fenntarthatósággal kapcsolatos szavak száma is több.

2.2.2. Leíró statisztika

A leíró statisztikai elemzést az IBM SPSS Statistics 29.0.1 szoftverével készítettem. A következő táblázatban az egyes cégek jelentéseiben lévő fenntarthatósággal kapcsolatos szavak átlaga, szórása, minimuma és maximuma látható.

4. táblázat
Negyedéves jelentésekben szereplő fenntarthatósággal kapcsolatos szavak leíró statisztikái

Megnevezés	Átlag	Szórás	Minimum	Maximum
BP	16	7	8	29
Eni	19	8	5	37
Galp Energia	13	7	0	31
MOL	6	3	2	11
Neste	22	17	6	61
OMV	8	3	4	13
PKN Orlen	36	21	7	75
Repsol	16	7	7	33
Shell	13	8	2	27
TotalEnergies	23	10	3	33
Összesen	17	9	4	35

Forrás: saját szerkesztés

Átlagosan a legtöbb fenntarthatósággal kapcsolatos szó a PKN Orlen negyedéves jelentéseiben található a legkevesebb pedig a MOL jelentéseiben. Ez a jelentések terjedelmével függ össze, míg a PKN Orlen jelentései átlagosan 70 oldal hosszúságúak, addig a MOL átlagosan egy 40 diából álló prezentációt tesz közé, ami viszonylag kevés szöveget tartalmaz. A legnagyobb szórás ugyancsak a PKN Orlen jelentéseiben van, ami annak köszönhető, hogy a vállalat 2020-as jelentéseiben még viszonylag kevés fenntarthatósággal kapcsolatos szó volt megtalálható, viszont a 2021-es és 2022-es években ez az érték jelentősen megnőtt, továbbá a második és negyedik negyedéves jelentésekben jelentősen több fenntarthatósággal kapcsolatos szó található, mint az első és harmadik negyedéves jelentésekben. Összességében megállapítható, hogy az európai olajvállalatok jelentéseiben egyre több fenntarthatósággal kapcsolatos szó van, ami arra utal, hogy a cégek évről évre egyre nagyobb figyelmet fordítanak a fenntarthatósággal kapcsolatos nyilatkozataikra.

5. táblázat
Vállalatok negyedéves árfolyamváltozásainak leíró statisztikái

Megnevezés	Átlag	Szórás	Minimum	Maximum
BP	3,5%	15,0%	-34,6%	30,3%
Eni	4,9%	17,8%	-24,7%	56,3%
Galp Energia	3,5%	12,7%	-23,5%	21,2%
MOL	2,5%	7,7%	-10,1%	16,1%

Neste	0,1%	16,5%	-25,9%	39,9%
OMV	4,2%	24,3%	-32,5%	82,7%
PKN Orlen	1,6%	17,1%	-26,4%	50,0%
Repsol	5,5%	24,4%	-31,1%	82,9%
Shell	4,1%	16,2%	-19,5%	44,3%
TotalEnergies	4,5%	12,9%	-19,1%	33,9%
Összesen	3,5%	15,0%	-34,6%	30,3%

Forrás: saját szerkesztés

Az olajvállalatok részvényárfolyamainak negyedéves emelkedése jelentősen megugrott az elmúlt négy évben, ami a COVID-19 pandémia okozta részvénypiaci esés miatti alacsony bázisnak köszönhető. Ezzel együtt az árfolyamok volatilitása is megnőtt, ezért láthatunk a táblázatban akár 24%-os szórást is egyes cégek esetében. Az elmúlt három évben két jelentős részvénypiaci esés is volt. Az első a COVID-19 pandémia, amikor az olajárak volt, hogy mínuszba fordultak, a második esést pedig a 2022 februárjában kirobbant orosz-ukrán háború okozta. Ugyanakkor az Oroszországra kivetett szankciók hatására magas energiaárak alakultak ki és bővültek a finomítói árak is, emiatt az olajvállalatok rég nem látott nyereséget könyvelhettek el 2022-ben. Ennek köszönhetően gyorsan felpattantak a tőzsdei árfolyamok is. Ezt a megnőtt volatilitást jól ábrázolja a minimum és maximum változások különbsége. Például az OMV esetében volt olyan negyedév, amikor 32,5%-ot esett a részvényárfolyam, de volt olyan negyedév is, amikor 82,7%-ot emelkedett.

2.2.3. Szavak közötti korreláció vizsgálata

A következő táblázatban a BP egyes negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak természetes eloszlása látható.

6. táblázat
BP negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak természetes eloszlása

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,32	0,473	0,38	0,382	0,382	0,46	0,498	0,421	0,294	0,392	0,251	0,416	0,388	0,474	0,429	0,39
Shapiro- Wilk p	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001

Forrás: saját szerkesztés

A Shapiro-Wilk teszt eredménye minden negyedéves jelentés esetén szignifikáns ($p < 0,001$), így kijelenthető, hogy a BP egyes negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak nem természetes eloszlásúak. A többi olajvállalat esetében is a Shapiro-Wilk teszt eredménye szignifikáns ($p < 0,001$), így ezen vállalatok egyes negyedéves

jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak sem természetes eloszlásúak. A maradék kilenc olajcég Shapiro-Wilk teszt eredménye az 1. számú mellékletben látható. Mivel az adatok nem természetes eloszlásúak, ezért a Spearman-féle korrelációt használtam. Ez a módszer a Pearson-féle korrelációval ellentétben nem csak a lineáris kapcsolatot vizsgálja, hanem azt is, hogy az értékek növekedése vagy csökkenése hogyan korrelál a másik változó rangsorával, ezáltal alkalmas nem normális eloszlású adathalmaz korrelációjának elemzésére.

A következő táblázatokban az elemzett vállalatok 2020. és 2023. közötti negyedéves jelentéseiben található fenntarthatósággal kapcsolatos szavak közötti korrelációt vizsgálok meg, amihez ugyancsak az IBM SPSS szoftverét használtam.

7. táblázat
BP negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	0,732**	1,000														
2020 Q3	0,524**	0,355*	1,000													
2020 Q4	0,722**	0,534**	0,745**	1,000												
2021 Q1	0,515**	0,529**	0,565**	0,570**	1,000											
2021 Q2	0,320*	0,703**	0,363*	0,367*	0,738**	1,000										
2021 Q3	0,681**	0,750**	0,614**	0,791**	0,797**	0,700**	1,000									
2021 Q4	0,665**	0,706**	0,531**	0,752**	0,729**	0,622**	0,939**	1,000								
2022 Q1	0,472**	0,503**	0,281	0,557**	0,517**	0,515**	0,753**	0,748**	1,000							
2022 Q2	0,442**	0,775**	0,139	0,370*	0,476**	0,771**	0,626**	0,602**	0,551**	1,000						
2022 Q3	0,549**	0,588**	0,125	0,417**	0,351*	0,417**	0,643**	0,658**	0,895**	0,639**	1,000					
2022 Q4	0,591**	0,562**	0,463**	0,702**	0,443**	0,428**	0,823**	0,800**	0,885**	0,456**	0,778**	1,000				
2023 Q1	0,223	0,296	0,432**	0,493**	0,273	0,316*	0,524**	0,500**	0,725**	0,334*	0,596**	0,638**	1,000			
2023 Q2	0,269	0,638**	0,473**	0,521**	0,522**	0,695**	0,669**	0,612**	0,576**	0,628**	0,455**	0,494**	0,785**	1,000		
2023 Q3	0,349*	0,371*	0,591**	0,627**	0,252	0,280	0,617**	0,578**	0,675**	0,284	0,544**	0,770**	0,761**	0,590**	1,000	
2023 Q4	0,382*	0,415**	0,466**	0,670**	0,299	0,335*	0,676**	0,631**	0,717**	0,313*	0,572**	0,827**	0,651**	0,509**	0,764**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

** Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

A BP esetében többségében 0,01 szignifikancia szint alatt egyre erősebb pozitív kapcsolat figyelhető meg az idő előrehaladtával a negyedéves jelentésekben lévő fenntarthatósággal kapcsolatos szavak között. Ez azt jelenti, hogy a jelentésekben használt fenntarthatósággal

kapcsolatos szavak korrelálnak egymással, így sok esetben ugyanazokat a szavakat használja a vállalat a negyedéves jelentéseiben. Ugyanakkor az is látható, hogy a vizsgált időszak elején és végén lévő jelentések között nincs vagy csak alacsony a korreláció, ami arra utal, hogy a jelentésekben használt fenntarthatósággal kapcsolatos szavak jelentősen eltérnek egymástól.

8. táblázat
Eni negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	0,572**	1,000														
2020 Q3	0,375*	0,747**	1,000													
2020 Q4	0,267	0,597**	0,735**	1,000												
2021 Q1	0,412**	0,631**	0,714**	0,525**	1,000											
2021 Q2	0,487**	0,694**	0,823**	0,644**	0,555**	1,000										
2021 Q3	0,521**	0,718**	0,821**	0,593**	0,667**	0,657**	1,000									
2021 Q4	0,458**	0,472**	0,481**	0,477**	0,294	0,649**	0,598**	1,000								
2022 Q1	0,619**	0,633**	0,763**	0,567**	0,568**	0,923**	0,721**	0,711**	1,000							
2022 Q2	0,439**	0,812**	0,988**	0,740**	0,676**	0,867**	0,822**	0,526**	0,794**	1,000						
2022 Q3	0,330*	0,738**	0,799**	0,587**	0,870**	0,699**	0,728**	0,419**	0,693**	0,775**	1,000					
2022 Q4	0,630**	0,660**	0,769**	0,571**	0,505**	0,785**	0,719**	0,480**	0,718**	0,797**	0,651**	1,000				
2023 Q1	0,510**	0,689**	0,846**	0,640**	0,611**	0,837**	0,791**	0,526**	0,780**	0,864**	0,722**	0,920**	1,000			
2023 Q2	0,703**	0,728**	0,683**	0,532**	0,551**	0,838**	0,651**	0,538**	0,773**	0,736**	0,697**	0,936**	0,835**	1,000		
2023 Q3	0,660**	0,827**	0,764**	0,577**	0,733**	0,748**	0,724**	0,449**	0,697**	0,792**	0,830**	0,857**	0,754**	0,922**	1,000	
2023 Q4	0,696**	0,727**	0,658**	0,475**	0,340*	0,651**	0,607**	0,422**	0,588**	0,716**	0,498**	0,908**	0,800**	0,841**	0,774**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

**. Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

A táblázat alapján megfigyelhető, hogy a fenntarthatósági szavak között általában többségében 0,01 szignifikancia szint alatt jelentős kapcsolat figyelhető meg az Eni negyedéves jelentéseiben található fenntarthatósággal kapcsolatos szavak között. Mindemellett a BP-vel ellentétben, az Eni esetében a vizsgált időszak eleji és végi jelentések között ugyanúgy korreláció áll fenn. Ez azt jelenti, hogy az Eni időszak eleji és végi jelentéseinek fenntarthatósággal kapcsolatos szóhasználata között nem fedezhető fel lényeges különbség.

9. táblázat
Galp Energia negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2020 Q2	-	1,000														
2020 Q3	-	0,733**	1,000													
2020 Q4	-	0,794**	0,910**	1,000												
2021 Q1	-	0,486**	0,610**	0,516**	1,000											
2021 Q2	-	0,465**	0,582**	0,680**	0,427**	1,000										
2021 Q3	-	0,465**	0,582**	0,680**	0,427**	1,000**	1,000									
2021 Q4	-	0,475**	0,585**	0,686**	0,436**	0,999**	0,999**	1,000								
2022 Q1	-	0,367*	0,351*	0,395**	0,354*	0,827**	0,827**	0,816**	1,000							
2022 Q2	-	0,316*	0,299	0,342*	0,302	0,729**	0,730**	0,719**	0,893**	1,000						
2022 Q3	-	0,245	0,250	0,304	0,274	0,592**	0,599**	0,584**	0,707**	0,656**	1,000					
2022 Q4	-	0,175	0,176	0,225	0,196	0,734**	0,734**	0,728**	0,877**	0,780**	0,806**	1,000				
2023 Q1	-	0,483**	0,465**	0,487**	0,175	0,438**	0,444**	0,431**	0,538**	0,494**	0,785**	0,623**	1,000			
2023 Q2	-	0,655**	0,596**	0,665**	0,348*	0,561**	0,567**	0,572**	0,453**	0,410**	0,680**	0,532**	0,869**	1,000		
2023 Q3	-	0,648**	0,591**	0,658**	0,342*	0,562**	0,567**	0,572**	0,460**	0,416**	0,688**	0,539**	0,877**	1,000**	1,000	
2023 Q4	-	0,397**	0,327*	0,370*	0,099	0,583**	0,583**	0,594**	0,459**	0,392*	0,308*	0,539**	0,530**	0,707**	0,707**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

**. Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

A Galp Energia 2020. első negyedéves jelentésében nem található fenntarthatósághoz kötődő szó, így a korrelációs vizsgálatot sem lehet elvégezni. A többi esetben, viszont nagyrészt 0,01 szignifikancia szint alatt jelentős kapcsolat figyelhető meg a negyedéves jelentésekben található fenntarthatósággal kapcsolatos szavak között. Ugyanakkor az is megállapítható, hogy az azonos években közzétett jelentések között magasabb a korreláció, azaz erősebb a kapcsolat, mint a másik három év negyedéves jelentéseivel. Ebből arra lehet következtetni, hogy azonos években többször használják ugyanazokat a szavakat, mint a másik három évben.

10. táblázat
MOL negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	0,607**	1,000														
2020 Q3	0,464**	0,569**	1,000													
2020 Q4	0,607**	0,500**	0,569**	1,000												
2021 Q1	0,711**	0,653**	0,455**	0,624**	1,000											
2021 Q2	0,464**	0,569**	0,641**	0,569**	0,733**	1,000										
2021 Q3	0,607**	0,725**	0,806**	0,724**	0,591**	0,806**	1,000									
2021 Q4	0,471**	0,577**	0,649**	0,577**	0,461**	0,649**	0,816**	1,000								
2022 Q1	0,368*	0,466**	0,839**	0,466**	0,363*	0,532**	0,679**	0,540**	1,000							
2022 Q2	0,385*	0,486**	0,862**	0,486**	0,381*	0,554**	0,707**	0,561**	0,714**	1,000						
2022 Q3	0,601**	0,486**	0,554**	0,743**	0,619**	0,554**	0,707**	0,563**	0,454**	0,472**	1,000					
2022 Q4	0,687**	0,624**	0,455**	0,392*	0,539**	0,455**	0,591**	0,766**	0,363*	0,381*	0,381*	1,000				
2023 Q1	0,604**	0,748**	0,547**	0,728**	0,655**	0,547**	0,697**	0,553**	0,444**	0,468**	0,465**	0,624**	1,000			
2023 Q2	0,464**	0,569**	0,641**	0,569**	0,455**	0,641**	0,806**	0,649**	0,532**	0,554**	0,554**	0,455**	0,547**	1,000		
2023 Q3	0,464**	0,569**	0,641**	0,569**	0,455**	0,641**	0,806**	0,649**	0,861**	0,554**	0,554**	0,455**	0,547**	0,641**	1,000	
2023 Q4	0,580**	0,720**	0,509**	0,445**	0,632**	0,509**	0,652**	0,838**	0,414**	0,433**	0,433**	0,888**	0,722**	0,509**	0,509**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

**. Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

Az adatok elemzése során megfigyelhető, hogy a MOL egyes negyedéves jelentéseiben használt fenntarthatósággal kapcsolatos szavak között többségében 0,01 szignifikancia szint alatt számos jelentős korreláció van. Azaz a vállalat ez irányú kommunikációjában a szóhasználat hasonlóan alakult a vizsgált időszak során.

11. táblázat
Neste negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	0,725**	1,000														
2020 Q3	0,794**	0,716**	1,000													
2020 Q4	0,835**	0,780**	0,670**	1,000												
2021 Q1	0,470**	0,802**	0,421**	0,702**	1,000											

2021 Q2	0,481**	0,804**	0,434**	0,711**	1,000**	1,000										
2021 Q3	0,137	0,515**	0,225	0,442**	0,734**	0,735**	1,000									
2021 Q4	0,577**	0,714**	0,465**	0,700**	0,678**	0,678**	0,644**	1,000								
2022 Q1	0,571**	0,315*	0,483**	0,616**	0,276	0,287	0,331*	0,404**	1,000							
2022 Q2	0,236	0,579**	0,154	0,485**	0,770**	0,769**	0,680**	0,713**	0,236	1,000						
2022 Q3	0,368*	0,715**	0,497**	0,659**	0,723**	0,724**	0,794**	0,652**	0,547**	0,669**	1,000					
2022 Q4	0,473**	0,633**	0,302	0,542**	0,493**	0,492**	0,589**	0,800**	0,462**	0,678**	0,563**	1,000				
2023 Q1	0,161	0,577**	0,247	0,314*	0,623**	0,622**	0,678**	0,436**	0,161	0,579**	0,655**	0,479**	1,000			
2023 Q2	0,503**	0,487**	0,197	0,389*	0,511**	0,510**	0,586**	0,624**	0,328*	0,461**	0,355*	0,690**	0,429**	1,000		
2023 Q3	0,370*	0,362*	0,239	0,270	0,392*	0,391*	0,428**	0,373*	0,392*	0,353*	0,416**	0,419**	0,493**	0,639**	1,000	
2023 Q4	0,519**	0,666**	0,420**	0,548**	0,559**	0,559**	0,547**	0,590**	0,252	0,386*	0,497**	0,608**	0,596**	0,628**	0,576**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

** .Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

A Neste esetében tizennégy esetben nincs kapcsolat két negyedéves jelentés. A többi esetben túlnyomó részt 0,01 szignifikancia szint alatt jelentős kapcsolat figyelhető meg az egyes negyedéves jelentésekben használt szavak között. A Neste jelentéseiben a fenntarthatósággal kapcsolatos szavak között általában közepes korrelációt látható, azaz a korreláció átlagosan 0,4-0,7 közötti értékeket vesz fel.

12. táblázat

OMV negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	1,000**	1,000														
2020 Q3	0,768**	0,768**	1,000													
2020 Q4	0,679**	0,679**	0,491**	1,000												
2021 Q1	0,679**	0,679**	0,491**	0,417**	1,000											
2021 Q2	0,747**	0,747**	0,552**	0,475**	0,665**	1,000										
2021 Q3	0,727**	0,727**	0,535**	0,461**	0,669**	0,999**	1,000									
2021 Q4	0,588**	0,588**	0,407**	0,827**	0,491**	0,785**	0,783**	1,000								
2022 Q1	0,768**	0,768**	0,568**	0,491**	0,685**	0,999**	0,998**	0,783**	1,000							
2022 Q2	0,599**	0,599**	0,419**	0,348*	0,509**	0,804**	0,802**	0,603**	0,802**	1,000						
2022 Q3	0,659**	0,659**	0,474**	0,401**	0,581**	0,904**	0,904**	0,695**	0,902**	0,719**	1,000					
2022 Q4	0,496**	0,496**	0,326*	0,408**	0,613**	0,734**	0,737**	0,664**	0,732**	0,727**	0,803**	1,000				

2023 Q1	0,601**	0,601**	0,421**	0,519**	0,548**	0,827**	0,830**	0,772**	0,828**	0,647**	0,913**	0,869**	1,000			
2023 Q2	0,698**	0,698**	0,506**	0,431**	0,611**	0,904**	0,904**	0,697**	0,906**	0,716**	0,996**	0,798**	0,913**	1,000		
2023 Q3	0,754**	0,754**	0,556**	0,479**	0,679**	0,998**	0,998**	0,781**	0,999**	0,804**	0,904**	0,734**	0,830**	0,904**	1,000	
2023 Q4	0,600**	0,600**	0,421**	0,351*	0,703**	0,839**	0,839**	0,634**	0,836**	0,658**	0,751**	0,718**	0,678**	0,748**	0,837**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

** .Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

Az OMV negyedéves jelentéseiben fellelhető fenntarthatósággal kapcsolatos szavak között is jelentős kapcsolat található többségében 0,01 szignifikancia szint mellett. Ugyanakkor sok esetben nagyon magas, 0,9-es érték feletti korreláció figyelhető meg, amely erős kapcsolatot jelent e jelentésekben található fenntarthatósággal kapcsolatos szavak között. Összességében az OMV esetében is megállapítható, hogy negyedéves jelentéseiben hasonló a fenntarthatósággal kapcsolatos szóhasználata.

13. táblázat

PKN Orlen negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	0,791**	1,000														
2020 Q3	0,781**	0,730**	1,000													
2020 Q4	0,616**	0,534**	0,805**	1,000												
2021 Q1	0,336*	0,254	0,428**	0,337*	1,000											
2021 Q2	0,346*	0,229	0,381*	0,502**	0,774**	1,000										
2021 Q3	0,425**	0,310*	0,469**	0,386*	0,855**	0,888**	1,000									
2021 Q4	0,377*	0,261	0,415**	0,439**	0,801**	0,951**	0,926**	1,000								
2022 Q1	0,032	-0,051	0,111	0,246	0,590**	0,548**	0,486**	0,593**	1,000							
2022 Q2	0,275	0,145	0,283	0,502**	0,637**	0,739**	0,605**	0,687**	0,797**	1,000						
2022 Q3	0,440**	0,278	0,411**	0,539**	0,654**	0,727**	0,631**	0,703**	0,761**	0,863**	1,000					
2022 Q4	0,400**	0,240	0,368*	0,533**	0,626**	0,725**	0,598**	0,661**	0,705**	0,929**	0,956**	1,000				
2023 Q1	0,118	0,129	0,165	0,288	0,651**	0,558**	0,517**	0,464**	0,505**	0,688**	0,612**	0,701**	1,000			
2023 Q2	0,195	0,180	0,299	0,504**	0,611**	0,670**	0,550**	0,603**	0,590**	0,800**	0,656**	0,721**	0,832**	1,000		
2023 Q3	0,097	0,099	0,140	0,334*	0,619**	0,623**	0,485**	0,538**	0,603**	0,765**	0,610**	0,689**	0,875**	0,938**	1,000	
2023 Q4	0,110	0,119	0,170	0,283	0,649**	0,551**	0,518**	0,465**	0,527**	0,694**	0,544**	0,629**	0,911**	0,895**	0,960**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

** .Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

A PKN Orlen esetében a legkevesebb a kapcsolat az egyes negyedéves jelentések között. Megfigyelhető, hogy a 2020-as negyedéves jelentések között van kapcsolat, viszont a többi jelentés és a 2020-as jelentések között viszonylag kevés kapcsolat található, különösen igaz ez a 2020 második negyedéves jelentésre. Ez abból adódik, hogy a PKN Orlen 2020-as jelentéseiben még kevés fenntarthatósággal kapcsolatos szó található a másik három vizsgált év negyedéves jelentéseivel összehasonlítva, emiatt nincs is kapcsolat közöttük. Mindemellett 2021., 2022. és 2023. évi negyedéves jelentések között 0,01 szignifikancia szint alatt markáns kapcsolat figyelhető meg.

14. táblázat
Repsol negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	0,526**	1,000														
2020 Q3	0,817**	0,662**	1,000													
2020 Q4	0,486**	0,738**	0,622**	1,000												
2021 Q1	0,496**	0,342*	0,639**	0,321*	1,000											
2021 Q2	0,446**	0,399**	0,458**	0,503**	0,674**	1,000										
2021 Q3	0,265	0,499**	0,370*	0,651**	0,316*	0,663**	1,000									
2021 Q4	0,391*	0,664**	0,528**	0,591**	0,647**	0,710**	0,600**	1,000								
2022 Q1	0,306*	0,368*	0,438**	0,500**	0,703**	0,705**	0,515**	0,721**	1,000							
2022 Q2	0,271	0,365*	0,252	0,486**	0,340*	0,524**	0,498**	0,559**	0,560**	1,000						
2022 Q3	0,370*	0,411**	0,480**	0,596**	0,424**	0,469**	0,616**	0,540**	0,482**	0,626**	1,000					
2022 Q4	0,446**	0,488**	0,561**	0,709**	0,503**	0,579**	0,733**	0,649**	0,591**	0,579**	0,862**	1,000				
2023 Q1	0,370*	0,411**	0,480**	0,601**	0,424**	0,474**	0,622**	0,544**	0,487**	0,470**	0,744**	0,870**	1,000			
2023 Q2	0,252	0,500**	0,354*	0,444**	0,496**	0,446**	0,462**	0,749**	0,485**	0,440**	0,574**	0,684**	0,579**	1,000		
2023 Q3	0,274	0,495**	0,380*	0,481**	0,515**	0,476**	0,500**	0,590**	0,513**	0,472**	0,826**	0,734**	0,623**	0,654**	1,000	
2023 Q4	0,306*	0,762**	0,412**	0,724**	0,358*	0,588**	0,739**	0,847**	0,584**	0,572**	0,661**	0,780**	0,667**	0,708**	0,720**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

** .Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

A korrelációs vizsgálatból kiderül, hogy a Repsol negyedéves jelentéseiben használt fenntarthatósággal kapcsolatos szavak között többségében 0,01 szignifikancia szint alatt közepes, 0,4-0,7 közötti értékű korreláció figyelhető meg. Egyedül a 2020. első negyedévi

jelentés esetében látható viszonylag kevés kapcsolat, ami arra utal, hogy ebben a jelentésben a szóhasználat általában eltért más jelentések szóhasználatától.

15. táblázat
Shell negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	0,703**	1,000														
2020 Q3	0,862**	0,589**	1,000													
2020 Q4	1,000**	0,703**	0,862**	1,000												
2021 Q1	0,785**	0,546**	0,670**	0,785**	1,000											
2021 Q2	0,618**	0,927**	0,702**	0,618**	0,483**	1,000										
2021 Q3	0,525**	0,539**	0,703**	0,525**	0,689**	0,701**	1,000									
2021 Q4	0,624**	0,652**	0,524**	0,624**	0,806**	0,617**	0,855**	1,000								
2022 Q1	0,495**	0,531**	0,408**	0,495**	0,652**	0,501**	0,716**	0,846**	1,000							
2022 Q2	0,623**	0,896**	0,516**	0,623**	0,490**	0,832**	0,531**	0,647**	0,760**	1,000						
2022 Q3	0,495**	0,531**	0,408**	0,495**	0,652**	0,501**	0,716**	0,846**	1,000**	0,760**	1,000					
2022 Q4	0,428**	0,456**	0,345*	0,428**	0,574**	0,426**	0,631**	0,753**	0,895**	0,666**	0,895**	1,000				
2023 Q1	0,607**	0,650**	0,509**	0,607**	0,785**	0,618**	0,854**	0,999**	0,847**	0,648**	0,847**	0,754**	1,000			
2023 Q2	0,647**	0,927**	0,537**	0,647**	0,512**	0,861**	0,546**	0,664**	0,542**	0,840**	0,542**	0,464**	0,665**	1,000		
2023 Q3	0,607**	0,650**	0,509**	0,607**	0,785**	0,618**	0,854**	0,999**	0,847**	0,648**	0,847**	0,754**	1,000**	0,665**	1,000	
2023 Q4	0,607**	0,650**	0,509**	0,607**	0,785**	0,618**	0,854**	0,999**	0,847**	0,648**	0,847**	0,754**	1,000**	0,665**	1,000**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

**. Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

A Shell negyedéves jelentéseiben található fenntarthatósággal kapcsolatos szavak közötti korrelációt vizsgálva kiderül, hogy a szavak között 0,01 szignifikancia alatt jelentős, néhány esetben erős kapcsolat figyelhető meg. Ebből arra lehet következtetni, hogy a vizsgált időszakban a Shell szóhasználata a fenntarthatósággal kapcsolatos kommunikációja során hasonló volt, nem tért el a jelentések között.

16. táblázat
TotalEnergies negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
korrelációs vizsgálata

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
2020 Q1	1,000															
2020 Q2	0,236	1,000														
2020 Q3	0,212	0,817**	1,000													
2020 Q4	-0,082	0,704**	0,735**	1,000												
2021 Q1	-0,091	0,628**	0,656**	0,725**	1,000											
2021 Q2	0,161	0,677**	0,692**	0,795**	0,690**	1,000										
2021 Q3	-0,091	0,628**	0,656**	0,725**	0,824**	0,842**	1,000									
2021 Q4	0,124	0,623**	0,639**	0,571**	0,774**	0,745**	0,774**	1,000								
2022 Q1	-0,108	0,526**	0,552**	0,620**	0,702**	0,572**	0,702**	0,762**	1,000							
2022 Q2	-0,082	0,509**	0,525**	0,594**	0,725**	0,611**	0,725**	0,686**	0,785**	1,000						
2022 Q3	-0,082	0,509**	0,524**	0,593**	0,724**	0,610**	0,725**	0,685**	0,784**	1,000**	1,000					
2022 Q4	-0,100	0,415**	0,429**	0,666**	0,603**	0,653**	0,604**	0,547**	0,786**	0,850**	0,850**	1,000				
2023 Q1	-0,108	0,533**	0,558**	0,782**	0,701**	0,719**	0,701**	0,766**	0,866**	0,784**	0,784**	0,779**	1,000			
2023 Q2	-0,099	0,411**	0,425**	0,493**	0,596**	0,621**	0,752**	0,537**	0,494**	0,672**	0,671**	0,552**	0,501**	1,000		
2023 Q3	-0,082	0,499**	0,516**	0,584**	0,531**	0,607**	0,721**	0,686**	0,785**	0,800**	0,799**	0,675**	0,809**	0,667**	1,000	
2023 Q4	-0,100	0,399**	0,414**	0,654**	0,592**	0,647**	0,593**	0,542**	0,637**	0,841**	0,840**	0,848**	0,775**	0,540**	0,663**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

**. Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

A TotalEnergies 2020 első negyedéves jelentése és a többi negyedéves jelentés között sehol sincs kapcsolat. Ez abból adódik, hogy a TotalEnergies 2020 első negyedéves jelentésében csak 3 fenntarthatósághoz kötődő szó található, míg átlagosan 23. A többi negyedéves jelentés között, viszont mindenhol 0,01 szignifikancia szint alatt markáns kapcsolat figyelhető meg a használt fenntarthatósággal kapcsolatos szavak között.

17. táblázat
Az elemzett vállalatok negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos
szavak korrelációs vizsgálata

	BP	Eni	Galp Energia	MOL	Neste	OMV	PKN Orlen	Repsol	Shell	TotalEnergies
BP	1,000									
Eni	0,658**	1,000								
Galp Energia	0,557**	0,539**	1,000							
MOL	0,462**	0,649**	0,453**	1,000						

Neste	0,517**	0,539**	0,347*	0,465**	1,000					
OMV	0,737**	0,652**	0,559**	0,583**	0,501**	1,000				
PKN Orlen	0,482**	0,622**	0,374*	0,467**	0,274	0,361*	1,000			
Repsol	0,482**	0,657**	0,279	0,338*	0,444**	0,504**	0,469**	1,000		
Shell	0,720**	0,542**	0,444**	0,619**	0,385*	0,665**	0,408**	0,421**	1,000	
TotalEnergies	0,670**	0,533**	0,520**	0,440**	0,403**	0,580**	0,484**	0,431**	0,607**	1,000

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

**.. Korreláció szignifikáns 0.01 szinten (két oldalú)

Forrás: saját szerkesztés

Az elemzett vállalatok összes negyedéves jelentésében használt fenntarthatósággal kapcsolatos szavak között többségében 0,01 szignifikancia szint alatt közepes, 0,4-0,7 közötti értékű korreláció figyelhető meg. Ez azt jelenti, hogy az elemzett vállalatok fenntarthatósággal kapcsolatos szóhasználata hasonló volt a vizsgált időszakban. Mindemellett látható magas korreláció is a BP és a Shell negyedéves jelentései között. Ugyanakkor vannak olyan vállalatok, amelyek negyedéves jelentései között nem található korreláció, ilyen a Repsol és a Galp Energia.

Összességében megállapítható, hogy korreláció van az elemzett vállalatok és negyedéves jelentéseik fenntarthatósághoz kötődő szóhasználata között. Azaz a vizsgált időszakban sok esetben ugyanazokat a szavakat használták a fenntarthatósággal kapcsolatos kommunikációjuk során.

2.2.4. A részvény árfolyamok és a fenntarthatósághoz kötődő szavak száma közötti kapcsolat vizsgálata

A következő táblázatban látható a használt fenntarthatósággal kapcsolatos szavak és a részvényárfolyamok Khí-négyzet próbája, ami azt mutatja meg, hogy van-e összefüggés közöttük.

18. táblázat

Negyedéves jelentésekben található fenntarthatósággal kapcsolatos szavak száma és a negyedéves árfolyamváltozások Khí-négyzet próbája

Megnevezés	Szabadságfok	Khí-négyzet
BP	165	0,265
Eni	165	0,265
Galp Energia	165	0,265
MOL	120	0,292
Neste	165	0,265

OMV	105	0,302
PKN Orlen	195	0,249
Repsol	180	0,257
Shell	150	0,273
TotalEnergies	150	0,273

Forrás: saját szerkesztés

A khí-négyzet próba eredménye a negyedéves árfolyamváltozásokra és a jelentésekben használt fenntarthatósággal kapcsolatos szavak számára egyik esetben sem érte el a 0,05 vagy az alatti kívánt értéket. Ez azt jelenti, hogy az árfolyamváltozások és a szavak száma között semmiféle összefüggés nem áll fenn, mivel mindegyik esetben jelentősen magasabb értékeket kaptunk, mint 0,05.

19. táblázat

Negyedéves jelentésekben található fenntarthatósággal kapcsolatos szavak száma és a negyedéves árfolyamváltozások közötti korreláció vizsgálata

Megnevezés	Pearson-féle korreláció	Spearman-féle korreláció
BP	-0,099 (p=0,716)	-0,065 (p=0,811)
Eni	0,238 (p=0,374)	0,507 (p=0,045)
Galp Energia	0,046 (p=0,864)	0,013 (p=0,961)
MOL	-0,268 (p=0,315)	-0,240 (p=0,371)
Neste	-0,193 (p=0,473)	-0,335 (p=0,204)
OMV	0,007 (p=0,980)	0,159 (p=0,556)
PKN Orlen	-0,013 (p=0,963)	0,112 (p=0,680)
Repsol	0,161 (p=0,550)	0,439 (p=0,089)
Shell	-0,128 (p=0,638)	-0,089 (p=0,743)
TotalEnergies	0,101 (p=0,709)	0,113 (p=0,678)

Forrás: saját szerkesztés

A negyedéves árfolyamváltozásokra és a negyedéves jelentésekben használt fenntarthatósággal kapcsolatos szavakra elvégeztem a Pearson- és Spearman-féle korrelációs vizsgálatot is. A kapott szignifikancia szintek egyik esetben sem érték el a kívánt 0,05 vagy az alatti értéket, így kijelenthető, hogy semmiféle kapcsolat nincs a negyedéves árfolyamváltozások és a jelentésekben lévő fenntarthatósággal kapcsolatos szavak száma között. Megfigyelhető, hogy sok esetben a negatív a korreláció, azaz például a fenntarthatósággal kapcsolatos szavak számának növekedése esetén enyhe árfolyamcsökkenés következik be az adott negyedév során. A kapott eredményt megerősíti a már elvégzett Khí-

négyzet próba eredménye is, ahol kiderült, hogy nincs összefüggés a szavak száma és az árfolyamváltozások között.

20. táblázat

Negyedéves jelentések teljesítménye az elemzői várakozásokhoz képest és a negyedéves árfolyamváltozások közötti korreláció vizsgálata

Megnevezés	Pearson-féle korreláció	Spearman-féle korreláció
BP	0,469 (p=0,067)	0,504 (p=0,046) *
Eni	0,200 (p=0,458)	0,219 (p=0,414)
Galp Energia	0,159 (p=0,556)	0,136 (p=0,617)
MOL	0,040 (p=0,883)	-0,044 (p=0,872)
Neste	0,024 (p=0,929)	-0,082 (p=0,763)
OMV	-0,467 (p=0,068)	-0,483 (p=0,058)
PKN Orlen	-0,417 (p=0,109)	-0,392 (p=0,133)
Repsol	-0,564 (p=0,023) *	-0,501 (p=0,048) *
Shell	-0,106 (p=0,697)	-0,157 (p=0,563)
TotalEnergies	-0,048 (p=0,861)	-0,052 (p=0,848)

*. Korreláció szignifikáns 0.05 szinten (két oldalú)

Forrás: saját szerkesztés

A fenti táblázatban az egyes társaságok negyedéves árfolyamváltozásainak és az egyes negyedéves jelentések elemzői konszenzusához mért alul-vagy felülteljesítés közötti korreláció látható. A táblázat alapján megállapítható, hogy az esetek többségében a vizsgált időszakban hiába teljesítették alul vagy felül az elemzői várakozásokat az egyes olajvállalatok ezek az ezt követő negyedévben nem voltak szignifikáns hatással a részvényeik árfolyamváltozásaira. A BP és a Repsol esetében mutat szignifikáns kapcsolatot a korrelációs vizsgálat, előbbi esetében pozitív, közepes erősséget, utóbbinál negatív, közepes erősséget.

21. táblázat

Fenntarthatósággal kapcsolatos szavak és a következő kereskedési nap árfolyam változásának korrelációja

Megnevezés	Pearson-féle korreláció	Spearman-féle korreláció
BP	0,092 (p=0,734)	0,173 (p=0,522)
Eni	0,149 (p=0,582)	0,006 (p=0,983)
Galp Energia	-0,368 (p=0,161)	-0,420 (p=0,105)
MOL	-0,038 (p=0,890)	0,010 (p=0,969)
Neste	-0,107 (p=0,692)	-0,127 (p=0,639)
OMV	-0,490 (p=0,054)	-0,466 (p=0,069)
PKN Orlen	-0,010 (p=0,969)	0,038 (p=0,888)

Repsol	-0,113 (p=0,676)	-0,075 (p=0,782)
Shell	0,228 (p=0,395)	0,226 (p=0,401)
TotalEnergies	-0,410 (p=0,114)	-0,361 (p=0,169)

Forrás: saját szerkesztés

A negyedéves jelentésekben szereplő fenntarthatósággal kapcsolatos szavak és a jelentést követő kereskedési nap árfolyamváltozása között ugyancsak nem figyelhető meg szignifikáns kapcsolat egyik vizsgált olaj cég esetében sem.

22. táblázat

Negyedéves jelentések teljesítménye az elemzői várakozásokhoz képest és a következő kereskedési nap árfolyam változásának korrelációja

Megnevezés	Pearson-féle korreláció	Spearman-féle korreláció
BP	-0,152 (p=0,575)	-0,112 (p=0,680)
Eni	0,398 (p=0,127)	0,336 (p=0,203)
Galp Energia	0,121 (p=0,656)	0,108 (p=0,689)
MOL	-0,027 (p=0,920)	-0,015 (p=0,957)
Neste	0,233 (p=0,386)	0,205 (p=0,446)
OMV	0,337 (p=0,201)	0,249 (p=0,353)
PKN Orlen	-0,143 (p=0,596)	-0,140 (p=0,605)
Repsol	0,249 (p=0,352)	0,313 (p=0,238)
Shell	0,075 (p=0,781)	0,157 (p=0,563)
TotalEnergies	0,393 (p=0,133)	0,295 (p=0,267)

Forrás: saját szerkesztés

A negyedéves jelentést követő kereskedési nap és az egyes negyedéves jelentések elemzői várakozásokhoz mért alul-vagy felülteljesítés között nem található szignifikáns kapcsolat. Tehát kijelenthető, hogy a jelentések elemzői konszenzushoz mért teljesítménye egyáltalán nem befolyásolja a másnapi árfolyamváltozást a jelen vizsgálat alapján.

2.3. KÖVETKEZTETÉSEK, JAVASLATOK

Ebben a részben következtetéseket vonok le az ismertett szakirodalom és a szekunder, primer kutatás alapján. A következtetések után javaslatokat teszek a vizsgált kérdés lehetséges további kutatásával kapcsolatban.

A természetes nyelvfeldolgozás és a mesterséges intelligencia felhasználása egyre több területen érezteti hatását. Ez igaz a tőzsdei elemzésekre is, ahol naponta rengeteg új információ keletkezik, ami hatással van a tőzsdei árfolyamokra. A természetes nyelvfeldolgozásnak és a

mesterséges intelligenciának köszönhetően lehetővé válik ennek a hatalmas adathalmaznak a feldolgozása. Ez pontosabb elemzéseket és gyorsabb hírfeldolgozást jelent, aminek köszönhetően gyorsabban átárazódnak a részvények az adott piaci helyzetnek megfelelően, azaz a piacok működése hatékonyabbá válik.

Szekunder kutatásom során arra a megállapításra jutottam, hogy azok az olajvállalatok, amelyek méretükhöz képest jól látható fenntartható tevékenységgel rendelkeznek, a piac magasabbra értékeli, azaz magasabb P/E rátával mozognak ezek a részvények. Mindemellett bemutattam, hogy ez a 10 olaj cég adja a világ szénhidrogén termelésének 6,13%-át, illetve az európai kőolaj-finomítás 59,8%-át. Ezekből az adatokból is látszik, hogy az elemzett vállalatok világszinten is, de európai viszonylatban még inkább a legnagyobb szereplők között vannak, így primer kutatásom eredményei akár kiterjeszthetők más európai olajvállalatokra is.

Dolgozatom primer kutatásában arra kerestem a választ, hogy van-e kapcsolat az egyes olajvállalatok negyedéves árfolyam változásai között, és a jelentésekben lévő fenntarthatósághoz kötődő szavak száma között. A negyedéves jelentésekben 2020 óta megnövekedtek a fenntarthatósággal kapcsolatos szavak. Az árfolyamváltozások a COVID-19 pandémia és az orosz-ukrán háború kitörése okozta általános tőzsdei esésnek köszönhetően jelentős kilengéseket mutattak. Az egyes vállalatok negyedéves jelentéseiben lévő fenntarthatósággal kapcsolatos szavak közötti korrelációs vizsgálat eredménye azt mutatta, hogy erős kapcsolat van a szavak között. Ez azt jelenti, hogy az egyes jelentésekben sok esetben ugyanazokat a fenntarthatósággal kapcsolatos szavakat használják hasonló mértékben. A Khí-négyzet próba eredménye szerint az árfolyamváltozások és a szavak száma között nincs összefüggés. A korrelációs vizsgálatok elvégzése alapján arra az eredményre jutottam, hogy a negyedéves árfolyamváltozások és a jelentésekben lévő fenntarthatósággal kapcsolatos szavak száma, illetve a jelentést követő kereskedés nap árfolyamváltozása és a jelentésekben lévő fenntarthatósággal kapcsolatos szavak száma között 5%-os szignifikancia szint mellett nincs kapcsolat. Ezt a Khí-négyzet próba eredménye is megerősítette. Ennek több oka lehet, például, hogy a dolgozatban használt szótár nem méri elég pontosan a negyedéves jelentésekben tett fenntarthatósághoz kötődő kijelentéseket vagy a jelentésekben még kevés fenntarthatósággal kapcsolatos információ található, hogy mérni lehessen vagy egyszerűen a befektetők szempontjából még nem elég hangsúlyos a fenntarthatóság, így nem figyelnek a negyedéves jelentésekben közzétett fenntarthatósággal kapcsolatos kijelentésekre, hanem jobban érdekli őket a vállalat pénzügyi teljesítménye. Ezenfelül a minta alacsony száma is magyarázatot adhat a szignifikáns kapcsolat hiányára, ezt a későbbiekben a minta növelésével lehetne orvosolni. Mindemellett elvégeztem az egyes negyedéves jelentések elemzői konszenzushoz mért

teljesítménye és a negyedéves árfolyamváltozások, illetve a jelentést követő kereskedés nap árfolyamváltozása közötti korrelációs vizsgálatot. Ennél a vizsgálatnál is azt az eredményt kaptam, hogy nincs kapcsolat a jelentések elemzői várakozásokhoz mért alul- vagy felülteljesítése és a negyedéves vagy jelentést követő napi árfolyamváltozás között. Ez abból adódhat, hogy az olajszektor nem rendelkezik nagy növekedési potenciállal, ezáltal jobban előrejelezhetőek a vállalati eredmények, így a közzétett jelentések nem okoznak akkora meglepetéseket, ami után jelentős árfolyamváltozások következnenek be.

A következő kutatásokban érdemes lehet a dolgozatban használt fenntarthatósággal kapcsolatos szótárat továbbfejleszteni kifejezetten az olajvállalatokra koncentrálva, így pontosabb képet kaphatnánk arról, hogy mennyit kommunikálnak ezen cégek a fenntarthatóságról a közzétett negyedéves jelentéseikben. További kutatási lehetőség még, hogy a természetes nyelvfeldolgozás helyett a fejlettebb mesterséges intelligenciát használjuk a jelentések elemzésére, így maga a szöveg értelmezése is megvalósulhatna a fenntarthatósággal kapcsolatos kulcsszavak keresése helyett.

ÖSSZEGZÉS

A természetes nyelvfeldolgozás (NLP) bemutatása során említést tettem a legfontosabb NLP módszerekről és algoritmusokról, a hangulatelemzésben és tőzsdei elemzésekben történő felhasználásáról. Az alkalmazási területei közül kiemelten a részvénypiaci volatilitás előrejelzését, illetve a válságok előrejelzését a jegybank elnöki beszédek elemzésével. Mindemellett ismertettem a közvélemény és politikai helyzet hatását a tőzsdére és a közösségi média üzenetek és tőzsdei árfolyamok kapcsolatát. Végül említést tettem a természetes nyelvfeldolgozás lehetséges jövőbeli továbbfejlesztéséről.

Szekunder kutatásomban ismertettem az elemzett vállalatok szénhidrogén kitermelését és kőolaj-finomításának mértékét, továbbá a különböző pénzügyi adataikat és mutatószámaikat. Kifejtettem, hogy azon cégek, amelyek méretükhöz képest rendelkeznek jelentős fenntartható tevékenységgel magasabb pénzügyi mutatószámokkal rendelkeznek, mint azok a vállalatok, amelyek nem. Ebből arra lehet következtetni, hogy a befektetők körében is megnőtt az elvárás, hogy azok a vállalatok, amelyekbe befektetik pénzüket azok fenntarthatóak is legyenek.

Primer kutatásomban bemutattam, hogy a vizsgált időszakban megnőtt a negyedéves jelentésekben lévő fenntarthatósággal kapcsolatos szavak száma. Ezt követően elvégeztem a

fenntarthatósággal kapcsolatos szavak számának és a negyedéves árfolyamváltozások leíró statisztikai elemzését. Korrelációs vizsgálatot végeztem az egyes negyedéves jelentésekben található fenntarthatósággal kapcsolatos szavak között. Ennek eredményeképpen azt a következtetést vontam le, hogy sok esetben ugyanazokat a fenntarthatósággal kapcsolatos szavakat használják az elemzett olajvállalatok a negyedéves jelentéseikben. Végül elkészítettem a fenntarthatósággal kapcsolatos szavak száma és a negyedéves tőzsdei árfolyamváltozások, illetve a fenntarthatósággal kapcsolatos szavak száma és a jelentést követő kereskedési nap árfolyamváltozása közötti korrelációs vizsgálatot. A kapott eredmény alapján, 5%-os szignifikancia szint mellett, valamint a jelenlegi információk ismeretében, nincs kapcsolat a szavak száma és az árfolyamváltozások között, azaz semmilyen mértékben nem befolyásolja a részvényárfolyamokat, ha az adott negyedéves jelentésben több vagy kevesebb fenntarthatósággal kapcsolatos szó található. Ezenfelül megvizsgáltam a negyedéves jelentések elemzői konszenzushoz mért alul- vagy felülteljesítésének és a negyedéves, jelentést követő kereskedés nap árfolyamváltozása közötti korrelációt. A kapott eredmény ez esetben sem mutatott 5%-os szignifikancia mellett jelentős kapcsolatot.

A természetes nyelvfeldolgozás és a mesterséges intelligencia megjelenésével új módszertanok alakultak ki a részvényárfolyamok előrejelzésére, amelyek a számadatok helyett a szöveges adatokat helyezik előtérbe és különféle nyelvi elemzési technikákat alkalmaznak a piaci trendek és befektetői hangulatok feltérképezésére. Dolgozatomban ezt az új módszert alkalmaztam egy átalakulóban lévő iparágra.

ÁBRAJEGYZÉK

1. ábra A szakirodalmi szűrés menete	4
2. ábra Fenntarthatósággal kapcsolatos szavak száma a negyedéves jelentésekben	21

TÁBLÁZATJEGYZÉK

1. táblázat Elemzett vállalatok bemutatása.....	17
2. táblázat Elemzett vállalatok 2023-as pénzügyi adatai milliárd dollárban	18
3. táblázat Elemzett vállalatok 2023-as pénzügyi mutatószámai	19

4. táblázat Negyedéves jelentésekben szereplő fenntarthatósággal kapcsolatos szavak leíró statisztikái.....	22
5. táblázat Vállalatok negyedéves árfolyamváltozásainak leíró statisztikái.....	22
6. táblázat BP negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak természetes eloszlása.....	23
7. táblázat BP negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	24
8. táblázat Eni negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	25
9. táblázat Galp Energia negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	26
10. táblázat MOL negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	27
11. táblázat Neste negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	27
12. táblázat OMV negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	28
13. táblázat PKN Orlen negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	29
14. táblázat Repsol negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	30
15. táblázat Shell negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	31
16. táblázat TotalEnergies negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata	32
17. táblázat Az elemzett vállalatok negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak korrelációs vizsgálata.....	32
18. táblázat Negyedéves jelentésekben található fenntarthatósággal kapcsolatos szavak száma és a negyedéves árfolyamváltozások Khí-négyzet próbája.....	33
19. táblázat Negyedéves jelentésekben található fenntarthatósággal kapcsolatos szavak száma és a negyedéves árfolyamváltozások közötti korreláció vizsgálata	34
20. táblázat Negyedéves jelentések teljesítménye az elemzői várakozásokhoz képest és a negyedéves árfolyamváltozások közötti korreláció vizsgálata	35

21. táblázat Fenntarthatósággal kapcsolatos szavak és a következő kereskedési nap árfolyam változásának korrelációja	35
22. táblázat Negyedéves jelentések teljesítménye az elemzői várakozásokhoz képest és a következő kereskedési nap árfolyam változásának korrelációja.....	36

IRODALOMJEGYZÉK

A. Yenter and A. Verma, “Deep CNN-LSTM with combined kernels from multiple branches for IMDB review sentiment analysis,” in Proc. IEEE 8th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON), Oct. 2017, pp. 540–546.

Agarwal, S., V. Y. S. Chen, and W. Zhang. 2016. “The Information Value of Credit Rating Action Reports: A Textual Analysis.” *Management Science* 6 (8): 2218–2240.

Alanyali, M., Moat, H. S., & Preis, T. (2013). Quantifying the relationship between financial news and the stock market. *Scientific Reports*, 3(1), 3578

Al-Hajieh, H., Redhead, K., & Rodgers, T. (2011). Investor sentiment and calendar anomaly effects: A case study of the impact of Ramadan on Islamic Middle Eastern markets. *Research in international Business and Finance*, 25(3), 345-356.

Atkins A, Niranjana M, Gerding E (2018) Financial news predicts stock market volatility better than close price. *J Financ Data Sci* 4(2):120–137

B. G. Malkiel, “The efficient market hypothesis and its critics,” *J. Econ. Perspect.*, vol. 17, no. 1, pp. 59–82, Feb. 2003.

Baccianella, S.; Esuli, A.; Sebastiani, F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, Paris, France, 17–23 May 2010; European Language Resources Association (ELRA): Valletta, Malta, 2010.

Behrendt, S., & Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking & Finance*, 96, 355–367

Bengio Y, Ducharme R, Vincent P (2003) A neural probabilistic language model. *J Mach Learn Res* 3:1137– 1155

Blei DM (2012) Probabilistic topic models. *Commun ACM* 55(4):77–84

- Brynjolfsson, E., D. Li, and L. R. Raymond. 2023. "Generative AI at Work." National Bureau of Economic Research Working Paper 31161
- Bühler K (1934) Sprachtheorie. Fischer, Jena
- Chan, S. W., & Chong, M. W. (2017). Sentiment analysis in financial texts. *Decision Support Systems*, 94, 53-64.
- Chen, B., Wu, Z., & Zhao, R. (2023). From fiction to fact: the growing role of generative AI in business and finance. *Journal of Chinese Economic and Business Studies*, 1-26.
- Chomsky N (1956) Three models for the description of language. *IRE Trans Inf Theory* 2(3):113–124.
- Day, M.-Y., & Lee, C.-C. (2016). Deep learning for financial sentiment analysis on finance news providers. In *Proceedings of the IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 1127–1134).
- De Arriba-Perez F, Garcí'a-Mendez S, Regueiro-Janeiro JA et al (2020) Detection of financial opportunities in micro-blogging data with a stacked classification system. *IEEE Access* 8:215,679– 215,690
- Delic V, Perić Z, Sećujski M et al (2019) Speech technology progress based on new machine learning paradigm. *Comput Intell Neurosci* 2019:1–19.
- Dowling, M., and B. Lucey. 2023. "ChatGpt for (Finance) Research: The Bananarama Conjecture." *Finance Research Letters* 53:103662.
- Farimani, S. A., Jahan, M. V., & Milani Fard, A. (2022). From Text Representation to Financial Market Prediction: A Literature Review. *Information*, 13(10), 466.
- Feuerriegel, S., & Gordon, J. (2018). Long-term stock index forecasting based on text mining of regulatory disclosures. *Decision Support Systems*, 112, 88-97.
- Forray, J. M., & Woodilla, J. (2005). Artefacts of management academe. *Time & Society*, 14(2–3), 323–339.
- Garcí'a-Mendez S, Ferná'andez-Gavilanes M, Juncal-Martí'nez J et al (2020) Identifying banking transaction descriptions via support vector machine short-text classification based on a specialized labelled corpus. *IEEE Access* 8:61,642–61,655.
- García-Méndez, S., de Arriba-Pérez, F., Barros-Vila, A., & González-Castaño, F. J. (2022). Detection of temporality at discourse level on financial news by combining Natural Language Processing and Machine Learning. *Expert Systems with Applications*, 197, 116648.
- García-Méndez, S., de Arriba-Pérez, F., Barros-Vila, A., González-Castaño, F. J., & Costa-Montenegro, E. (2023). Automatic detection of relevant information, predictions and forecasts

in financial news through topic modelling with Latent Dirichlet Allocation. *Applied Intelligence*, 1-19.

Gibbons, M. R., & Hess, P. (1981). Day of the week effects and asset returns. *Journal of business*, 579-596.

Hagenau, M., Hauser, M., Liebmann, M., & Neumann, D. (2013, January). Reading all the news at the same time: Predicting mid-term stock price developments based on news momentum. In 2013 46th Hawaii International Conference on System Sciences (pp. 1279-1288).

Hansen, A. L., and S. Kazinnik. 2023. Can ChatGpt Decipher FedSpeak?

Hassan, T. A., S. Hollander, L. V. Lent, and A. Tahoun. 2019. "Firm-Level Political Risk: Measurement and Effects." *The Quarterly Journal of Economics* 134 (4): 2135–2202.

Herwartz, H. (2017). Stock return prediction under GARCH—An empirical assessment. *International Journal of Forecasting*, 33(3), 569-580.

Herzenstein, M., S. Sonenshein, and U. M. Dholakia. 2011. "Tell Me a Good Story and I May Lend You Money: The Role of Narratives in Peer-To-Peer Lending Decisions." *Journal of Marketing Research* 48 (SPL): S138–S149.

<https://companiesmarketcap.com/oil-gas/largest-oil-and-gas-companies-by-market-cap/>
<https://finbox.com/>

Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & operations research*, 32(10), 2513-2522.

K. Mishev, A. Gjorgjevikj, I. Vodenska, L. Chitkushev, W. Souma, and D. Trajanov, "Forecasting corporate revenue by using deep-learning methodologies," in *Proc. Int. Conf. Control, Artif. Intell., Robot. Optim. (ICCAIRO)*, May 2019, pp. 115–120

Kamal, S., Sharma, S., Kumar, V., Alshazly, H., Hussein, H. S., & Martinetz, T. (2022). Trading stocks based on financial news using attention mechanism. *Mathematics*, 10(12), 2001.

Khan, W., Malik, U., Ghazanfar, M. A., Azam, M. A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. *Soft Computing*, 24, 11019-11043.

Klein, F. C., & Prestbo, J. A. (1974). *News and the Market*. H. Regnery Company.

Korangi, K., C. Mues, and C. Bravo. 2023. "A Transformer-Based Model for Default Prediction in Mid-Cap Corporate Markets." *European Journal of Operational Research* 308 (1): 306–320.

Krutilla, Z., & Kovari, A. (2022). Géppel történő természetes nyelvfeldolgozás üzleti alkalmazása. *Journal of Applied Technical and Educational Sciences*, 12(4), 327-327.

- Lakonishok, J., & Levi, M. (1982). Weekend effects on stock returns: a note. *The Journal of Finance*, 37(3), 883-889.
- Liu, Q., Lee, W. S., Huang, M., & Wu, Q. (2023). Synergy between stock prices and investor sentiment in social media. *Borsa Istanbul Review*, 23(1), 76-92.
- Liu, Q., Wang, X., & Du, Y. (2022). The weekly cycle of investor sentiment and the holiday effect--An empirical study of Chinese stock market based on natural language processing. *Heliyon*, 8(12).
- LOUGHRAN, T.; MCDONALD, B. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *J. Financ.* 2011, 66, 35–65.
- Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE access*, 8, 131662-131682.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), 264-269.
- Németh, R., Katona, E. R., & Kmetty, Z. (2020). Az automatizált szövegelemzés perspektívája a társadalomtudományokban. *Szociológiai Szemle*, 30(1), 44-62.
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53(4), 3007-3057.
- Oliveira, N.; Cortez, P.; Areal, N. Stock market sentiment lexicon acquisition using microblogging data and statistical measures. *Decis. Support Syst.* 2016, 85, 62–73.
- Oztekin, A., Kizilaslan, R., Freund, S., & Iseri, A. (2016). A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal of Operational Research*, 253(3), 697-710.
- Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, 33(6), 497-505.
- Park, J., & Cho, S. (2023). Incorporation of company-related factual knowledge into pre-trained language models for stock-related spam tweet filtering. *Expert Systems with Applications*, 234, 121021.
- Park, J., Lee, H. J., & Cho, S. (2023). Hot topic detection in central bankers' speeches. *Expert Systems with Applications*, 120563.

Pencle, N., & Mălăescu, I. (2016). What's in the words? Development and validation of a multidimensional dictionary for CSR and application using prospectuses. *Journal of Emerging Technologies in Accounting*, 13(2), 109-127.

Rickett, L. K. (2016). Do financial blogs serve an infomediary role in capital markets? *American Journal of Business*, 31(1), 17–40.

Rounaghi, M. M., & Zadeh, F. N. (2016). Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: monthly and yearly forecasting of time series stock returns using ARMA model. *Physica A: Statistical Mechanics and its Applications*, 456, 10-21.

Rustam F, Reshi AA, Mehmood A et al (2020) COVID-19 future forecasting using supervised machine learning models. *IEEE Access* 8:101,489–101,499.

S. F. Crone and C. Koeppel, “Predicting exchange rates with sentiment indicators: An empirical evaluation using text mining and multilayer perceptrons,” in *Proc. IEEE Conf. Comput. Intell. Financial Eng. Econ. (CIFEr)*, Mar. 2014, pp. 114–121

S. Zubair, K.J. Cios, Extracting news sentiment and establishing its relationship with the S&P 500 index, *Hawaii International Conference on System Sciences*, IEEE, 2015, pp. 969–975.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2), 1-19.

Schumann, J. H. (1987). The expression of temporality in basilean speech. *Studies in Second Language Acquisition*, 9(1), 21–41.

Siering, M. (2013). Investigating the impact of media sentiment and investor attention on financial markets. In *Enterprise Applications and Services in the Finance Industry: 6th International Workshop, FinanceCom 2012, Barcelona, Spain, June 10, 2012. Revised Papers* 6 (pp. 3-19). Springer Berlin Heidelberg.

Solorio-Fernandez S, Carrasco-Ochoa JA, Martínez-Trinidad JF (2020) A review of unsupervised feature selection methods. *Artif Intell Rev* 53(2):907–948.

Sowa JF (1987) Semantic networks. In: Shapiro SC (eds) *Encyclopedia of artificial intelligence*. Wiley, pp 1011–1024

Stevenson, M., C. Mues, and C. Bravo. 2021. “The Value of Text for Small Business Default Prediction: A Deep Learning Approach.” *European Journal of Operational Research* 295 (2): 758–771.

Swathi, T., Kasiviswanath, N., & Rao, A. A. (2022). An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis. *Applied Intelligence*, 52(12), 13675-13688.

Taimur M, Khan S (2015) Impact of political and catastrophic events on stock returns. *VFAST Trans Edu Soc Sci* 6(1):21–32

Teng, C. C., & Liu, V. W. (2013). The pre-holiday effect and positive emotion in the Taiwan Stock Market, 1971–2011. *Investment Analysts Journal*, 42(77), 35-43.

Vargas, M. R., de Lima, B. S. L. P., & Evsukoff, A. G. (2017). Deep learning for stock market prediction from financial news articles. In *Proceedings of the IEEE international conference on computational intelligence and virtual environments for measurement systems and applications* (pp. 60–65).

W. Souma, I. Vodenska, and H. Aoyama, “Enhanced news sentiment analysis using deep learning methods,” *J. Comput. Social Sci.*, vol. 2, no. 1, pp. 33–46, Jan. 2019.

Wang, H.; Lu, S.; Zhao, J. Aggregating multiple types of complex data in stock market prediction: A model-independent framework. *Knowl.-Based Syst.* 2019, 164, 193–204.

Wang, J. J., Wang, J. Z., Zhang, Z. G., & Guo, S. P. (2012). Stock index forecasting based on a hybrid model. *Omega*, 40(6), 758-766.

Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: a survey. *Artificial Intelligence Review*, 50(1), 49-73.

Yang, K. C., and F. Menczer. 2023. “Large language models can rate news outlet credibility.”

Yang, L.; Xu, Y.; Ng, J.; Dong, R. Leveraging BERT to improve the FEARS index for stock forecasting. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, Macao, China, 12 August 2019; ACL: Stroudsburg, PA, USA, 2019.

MELLÉKLETEK

1. sz. melléklet

Eni negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak természetes eloszlása

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,388	0,297	0,477	0,504	0,339	0,341	0,459	0,482	0,456	0,441	0,39	0,383	0,421	0,336	0,448	0,321
Shapiro- Wilk p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**Galp Energia negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
természetes eloszlása**

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	n.a.	0,281	0,302	0,327	0,316	0,346	0,321	0,248	0,303	0,294	0,21	0,262	0,283	0,322	0,303	0,335
Shapiro- Wilk p	n.a.	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**MOL negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
természetes eloszlása**

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,411	0,336	0,284	0,295	0,388	0,284	0,222	0,283	0,318	0,339	0,339	0,388	0,326	0,284	0,284	0,336
Shapiro- Wilk p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**Neste negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
természetes eloszlása**

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,35	0,411	0,235	0,363	0,28	0,339	0,355	0,414	0,382	0,373	0,388	0,453	0,203	0,25	0,324	0,436
Shapiro- Wilk p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**OMV negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
természetes eloszlása**

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,283	0,283	0,388	0,427	0,427	0,384	0,382	0,442	0,388	0,459	0,431	0,533	0,469	0,431	0,379	0,47
Shapiro- Wilk p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**PKN Orlen negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
természetes eloszlása**

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,384	0,5	0,403	0,481	0,381	0,338	0,279	0,307	0,46	0,429	0,519	0,572	0,498	0,527	0,528	0,508
Shapiro- Wilk p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**Repsol negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
természetes eloszlása**

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,310	0,346	0,276	0,368	0,235	0,404	0,32	0,427	0,315	0,467	0,317	0,269	0,283	0,267	0,32	0,39
Shapiro- Wilk p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**Shell negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
természetes eloszlása**

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,283	0,344	0,276	0,235	0,222	0,441	0,335	0,284	0,256	0,347	0,23	0,23	0,182	0,29	0,171	0,182
Shapiro- Wilk p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**TotalEnergies negyedéves jelentéseiben szereplő fenntarthatósággal kapcsolatos szavak
természetes eloszlása**

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4
N	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42
Shapiro- Wilk W	0,219	0,417	0,424	0,333	0,299	0,269	0,244	0,305	0,22	0,216	0,198	0,209	0,255	0,241	0,218	0,197
Shapiro- Wilk p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001