

TARTU ÜLIKOOL
Arvutiteaduse instituut
Infotehnoloogia mitteinformaatikutele õppekava

Anne-Liis Rämson
Pressinõukogule esitatud kaebuste otsuste
ennustamine masinõppe abil
Magistritöö (15 EAP)

Juhendaja Kairit Sirts, PhD

Tartu 2022

Pressinõukogule esitatud kaebuste otsuste ennustamine masinõppe abil

Lühikokkuvõte:

Käesolevas töös käsitletakse aastatel 2001-2021 Pressinõukogule esitatud kaebusi ja kaebustele vastavaid meediatekste. Töö eesmärkideks on anda statistiline ülevaade Pressinõukogule esitatud kaebustest ja Pressinõukogu otsustes nimetatud eetikakoodeksi punktide mainimistest, rakendada klassifitseerimismeetodeid kaebustele vastavatele meediatekstidele ning leida klassifitseeriv mudel, mis eristaks õigeksmõistva ja tauniva otsuse saanud meediatekste. Töö teoreetilises osas antakse ülevaade tekstikaevest, klassifitseerimismudelitest (logistiline regressioon, tugivektorklassifitseerija, fastText) ja klassifitseerimismudelite hindamismõõdikutest. Kaebuste analüüsimisel selgus, et Eesti suuremad väljaanded jagunevad eetikakoodeksi punktide mainimiste osas kahte gruppi. Leiti kolm suurt ajakirjandusväljaannet, mille kohta on esitatud kaebustes enam mainimisi saanud koodeksipunkt 4.2 ning kolm väljaannet, mille artiklite puhul on kõige enam mainimisi saanud koodeksipunkt 1.4. Taunivaid otsuseid prognoosis kõige paremini fastText klassifitseerija lemmatiseeritud tekstidel.

Võtmesõnad: Pressinõukogu, tekstikaeve, ajakirjandustekstide klassifitseerimine

CERCS: P176, Tehisintellekt

Predicting the outcome of the complaints submitted to the Press Council using machine learning

Abstract:

This thesis discusses complaints submitted to the Press Council between 2001 and 2021 and media texts in accordance with the complaints. The aim of this thesis is to provide statistical overview about the complaints lodged with the Press Council and mentioning the elements of the Code of Ethics in the decisions appointed by the Press Council, apply the methods of classification to media texts in line with the complaints and to identify a classification model that would distinguish acquittal and reprehensible media texts. The theoretical part of this thesis gives an overview about text mining, classification models (logistic regression, Support Vector Machine, fastText) and evaluation methods for classification. The analysis of complaints revealed that there were two main groups of the largest media outlets in Estonia with regard to mentioning the key elements of the Code of Ethics. It was found that three media outlets received more mentions about section 4.2 of the Code of Ethics and three media outlets received most mentions about section 1.4. Reprehensible decisions were best predicted by the fastText classifier on lemmatized texts.

Keywords: Estonian Press Council, text mining, classification of press texts

CERCS: P176, Artificial intelligence

Sisukord

1. Sissejuhatus	4
2. Ülevaade ajakirjanduse eneseregulatsioonist ja meediakriitikast.....	6
3. Tekstikaeve ja klassifitseerimismudelid	9
3.1 Tekstikaeve.....	9
3.2 Tekstide representeerimine	10
3.3 Klassifitseerimismudelid.....	11
3.3.1 Logistiline regressioon.....	11
3.3.2 Tugivektorklassifitseerija.....	12
3.3.3 Facebook fastText mudel	13
3.4 Mudelite hindamise meetodid	14
4. Pressinõukogu kaebuste kogumine, töötlemine, ülevaade andmetest	16
4.1 Kaebuste kogumine ja ettevalmistamine analüüsiks	16
4.2 Statistiline ülevaade kaebustest	18
4.3 Kaebuste linkide kogumine, tekstikogu loomine	24
5. Tekstide klassifitseerimine ja analüüs	27
5.1 Eksperimendi kirjeldus.....	27
5.2 Tulemused	29
5.3 Diskussioon, järeldused.....	31
6. Kokkuvõte	34
Kasutatud kirjandus	36
Lisad.....	37
I. Eesti ajakirjanduseetika koodeks	37
II. Tabelid.....	39
III. Illustreerivad joonised	46
IV. Litsents	47

1. Sissejuhatus

Tekstiliste andmete uurimine on saanud laiemat tähenduse andmekaeve meetodite arenguga. Struktureerimata andmete analüüsi toetavad masinõppemeetodid, mis oskavad tekstikujul olevatest andmetest leida seaduspärasusi, edastada infot tekstidesse peidetud mustritest. Käesolevas töös uuritakse ajakirjanduslikes väljaannetes ilmunud meediatekste.

Ajakirjanduse tulevikuvaates võiks olla oskus luua teksti, kus tekstis esitatud faktid ja arvamused on selgelt eristatavad. Meediatekstit võiksid olla neutraalsed ja tasakaalustatud. Ajakirjanikud vajaksid infokorratuse ajastul suuremat infokirjaoskust ja teadusvaldkondade tundmist. Kriitiline oskus hinnata uudiste faktipõhisust ning objektiivne sündmuste kajastus meediatekstides pakub võimaluse luua kvaliteetset ajakirjandust (Kikas & Himma-Kadakas, 2021). Kontrollitud faktidega erapooletud ja tõesed meediatekstit on pressivabaduse aluspõhimõtteks. Ajakirjaniku tööks on kirjeldada sündmusi ja olustikku objektiivselt, sõltumatult ja ühtlasi ka jälgida eetika- ja moraalinorme.

Pressinõukogu ülesandeks on olla vahelüli ajakirjaniku ja ilmunud meediatekstis kajastuse saanud füüsilise või juriidilise isiku vahel. Pressinõukogu tööks on analüüsida, kas ajakirjanduslikus tekstis esitatu on kooskõlas eetikakoodeksi punktidega, kas väited on faktipõhised, kas kõikidele osapooltele on antud sõna, kas on jälgitud moraalinorme. Pressinõukogu toimib meediatekstide jaoks filtrina, ajakirjanduslik tekst läbib analüüsi ja kontrolli ning tulemuseks on tauniv või õigeksmõistev otsus.

Käesolevas töös kirjeldatakse Pressinõukogule esitatud kaebusi, tuuakse välja sagedasemad eetikakoodeksi punktide mainimised, vaadeldakse Pressinõukogu analüüsi läbinud meediatekste, otsitakse seaduspärasusi, mis eristaksid tauniva ja õigeksmõistva otsuse saavaid tekste. Piisava eristusvõimega mudeli leidumise korral oleks see üheks tööriistaks, mille abil kontrollida publitseeritava teksti korrektsust. Meedia eneseregulatsiooni ja eetikakoodeksi punktide järgimist on varasemalt uuritud Eva Kübara (2006) ja Lenel Karu (2013) üliõpilastöodes. Töö autorile teadaolevalt ei ole Pressinõukogu otsustega seonduvaid meediatekste seni analüüsitud masinõppemeetoditega.

Uurimuse eesmärgid

- Anda statistiline ülevaade Pressinõukogule esitatud kaebustest ja otsustes mainitud eetikakoodeksi punktidest.
- Rakendada klassifitseerimismeetodeid kaebustele vastavatele õigeksmõistva või tauniva otsuse saanud meediatekstidele.

- Võrrelda erinevaid klassifitseerimismudeleid nende täpsuse osas.

Uurimisküsimused

- Milliseid eetikakoodeksi punkte mainitakse taunivates otsustes? Kas väljaannete lõikes on erinevusi tauniva otsuse saanud eetikakoodeksipunktide mainimistes?
- Kuidas ja kui hästi eristuvad tauniva ja õigeksmõistva otsuse saanud meediatekstid? Kas meediatekstide sõnapõhise klassifitseerimise korral on võimalik tuvastada tauniva otsuse saanud teksti?

Käesoleva töö ülesehitus

Töö teises peatükis tehakse ülevaade Eesti ajakirjanduse eneseregulatsioonivõimalustest, kirjeldatakse Pressinõukogu kui eneseregulatsiooniorgani tööd ning antakse lühike ülevaade meediakriitikaga tegelevatest institutsioonidest. Kolmandas peatükis on esitatud tekstide klassifitseerimisega seonduv teoreetiline osa, antakse ülevaade tekstikaevest, tekstide representeerimisvõimalustest ning kirjeldatakse lähemalt kolme antud töös rakendatud klassifitseerimismeetodit. Neljandas peatükis antakse statistiline ülevaade Pressinõukogule esitatud kaebustest ja kaebustele vastava tekstikogu loomisest leitud veebilinkide abil. Statistilises osas kirjeldatakse Pressinõukogule esitatud kaebuste ajalist jaotust, kaebuste jagunemist väljaannete vahel, eetikakoodeksi punktide mainimisi, eetikakoodeksi punktide mainimisi suuremate väljaannete lõikes, kaebuste jagunemist õigeksmõistvate ja taunivate kaebuste vahel. Viiendas peatükis kirjeldatakse kaebustele vastavate meediatekstide klassifitseerimist kolme erineva klassifitseerimismeetodiga, antakse ülevaade mudelite tulemustest ja järeldustest. Kuuendas peatükis tehakse töö tulemustest kokkuvõte, kirjeldatakse töö koostamisel tekkinud probleeme ning võimalusi töö edasiarendamiseks.

2. Ülevaade ajakirjanduse eneseregulatsioonist ja meediakriitikast

Pressivabadus on demokraatliku elukorralduse aluspõhimõte. Ajakirjanduselt oodatakse erapooletult kajastatud tõest ja faktidega kontrollitud informatsiooni. Lisaks pressivabadusele on oluline, et ajakirjanduses toimiks ka meediajärelevalve. Üheks meediajärelevalve võimaluseks on ajakirjanduse eneseregulatsioon, mis toetub kokkulepitud normide kogumikele (Kübar, 2006).

Eestis on ajakirjanduse eneseregulatsiooni aluseks ajakirjanduseetika koodeks, mille järgimine aitab ajakirjanikul publitseeritavates tekstides säilitada väärtusi ja moraalseid käitumisprintsippe. Ajakirjaniku teadlikkus eetika- ja moraalinormidest on aluseks, et tagada lugejale kvaliteetset ajakirjanduslikku teksti. Vähesese töökogemusega ajakirjanike jaoks ei ole ajakirjanduseetika koodeks teejuhiks, kuid eetilist mõtlemist peetakse oluliseks (Karu, 2013).

Meedias levivat infokorruptust ehk infosegadust võib jagada kolme rühma. Esimesse rühma kuulub eksitav info ehk väärinformatsioon (*misinformation*), mille levitamise eesmärk ei ole pahatahtlik, sellega ei soovita kahju tekitada. Desinformatsiooni (*disinformation*) levitatakse teadlikult ja tahtlikult kahju tekitamise eesmärgil. Kolmandasse rühma kuulub vaenulik info (*malinformation*), mille eesmärk on isik, organisatsiooni või riigi kahjustamine (Wardle & Derakhsan, 2017).

Pressinõukogu on loodud 2002. aastal eneseregulatsiooni organina Eesti Meediaettevõtete Liidu (tollase nimega Eesti Ajalehtede Liit) poolt. Pressinõukogusse kuulub kümme liiget, lisaks ajakirjanikele on liikmeid ka meediavälistest sektoritest. Pressinõukogu poole võib pöörduda kaebusega ajakirjanduses ilmunud materjalide kohta iga füüsiline või juriidiline isik. Ajakirjanduskonfliktide aruteludes arvestab Pressinõukogu ühelt poolt ajakirjanduse pressivabadust, teisest poolt arvestatakse isikute, ettevõtete või institutsioonide õigustatud ootusi ja huve seoses ajakirjandusega. Pressinõukogul on õigus teha otsus kaebuse menetlusse võtmise kohta. Menetlusse võetud kaebuste otsused on taunivad või õigeksmõistvad. Pressinõukogu otsused on avalikud, mis avaldatakse Eesti Meediaettevõtete Liidu koduleheküljel. Tauniva otsuse saanud väljaanne avaldab otsuse muutmata kujul 10 päeva jooksul (Pressinõukogu, 2022).

Avaliku Sõna Nõukogu (ASN) loodi 1991. aastal. Avaliku Sõna Nõukogu arutab avaliku meedia kohta esitatud kaebusi ja juhib ka ise tähelepanu meediaeetilistele probleemidele. ASN esitab oma meediaeetikaalased seisukohad kas vabastavate või taunivate lahenditena. Avaliku Sõna Nõukogu tegutses aastatel 1991–2001 eeskätt ajakirjanduse eneseregulatsiooni organina.

Erimeelsuste tõttu sai Avaliku Sõna Nõukogu teatavas mõttes järglaseks Pressinõukogu. Kogu tegutsemise aja jooksul on Avaliku Sõna Nõukogu käsitlenud üle 700 kaebuse (Avaliku Sõna Nõukogu, 2021).

Meediakriitikaga tegelevad Eestis mitmed portaalid. Propastop.org blogi tegeleb Eesti-vastaste manipulatsioonide, valede ja propaganda paljastamisega, eesmärgiks on kaasa aidata Eesti informatsiooni turvalisuse säilimisele. Valdavalt kajastatakse ja paljastatakse Venemaaga seotud väärted, kuid vajadusel avalikustatakse ka teisi riike või organisatsioone, mis soovivad mõjutada Eestis toimuvat (Propastop, 2022).

Ekspress Meedia uuriva toimetuse juures tegutsev Eesti Päevalehe Faktikontroll jälgib Eesti avalikus informatsioonis avaldatud väidete täpsust. Faktikontrolli tehes jälgitakse väite faktipõhisust, faktikontrolli läbipaistvust ja erapooletust. Faktikontrolli otsused on järgmised – tõsi, pigem tõsi, nii ja naa, eksitav, pigem vale, vale (Faktikontroll, 2022). Meediakriitikaga tegelevad veel mitmed portaalid ja tele-raadiosaated – ERR meediakriitika portaal, Vikerraadio „Meediatund“, Kuku Raadio „Olukorrast ajakirjanduses“, Objektiiv portaali meediakriitika.

Globaalne Väärinformatsiooni Indeks (*The Global Disinformation Index – GDI*) pakub riskihinnanguid maailma uudisportaalidele. GDI on Ühendkuningriigis asuv meediat uuriv apoliitiline mittetulundusühing. Eestis uuriti uudisportaalide võimalikku väärinfo levitamist 2020. aasta uuringus. GDI riskihinne koondub neljast sambast: uudiste struktuur, sisu, portaali funktsionaalsus, avalik arvamus. Neljale sambale toetudes antakse hinnang uudisportaaali võimalikule väärinfo levitamisele. Uuringu tulemustest selgus, et eksitava sisuga ajakirjanduslikke tekste kohtab Eesti online uudisportaalides harva. Uuringusse oli kaasatud 33 uudisportaaali. Madalaim väärinformatsiooni levik on antud uuringu põhjal Eesti Rahvusringhäälingu eesti- ja venekeelsetel lehekülgedel ning Äripäeva majandusuudiste portaalil. Kõrgeim väärinfo levimise riskitase omistati uuringus valdavalt peavoolust väljajäävatele uudisportaalidele. Suurem osa valimist jäi keskmise riski tasemele, need on regionaalsed ja üleriigilised uudisportaalid. Viimase kategooria uudisportaalidel ei ole selgeid reegleid kuidas toimetatakse ajakirjanduslikke lugusid, kuidas on tagatud ajakirjanduslik sõltumatus ja kuidas suheldakse lugejatega (GDI, 2020).

Piirideta Ajakirjanikud on rahvusvaheline organisatsioon, mis on asutatud pressivabaduse kaitseks. Kord aastas koostatakse riikide kaupa maailma pressivabaduse indeks. Indeksi koostamisel küsitletakse kohalikke ajakirjanikke. Indeks koostatakse viie aspekti põhjal:

poliitiline kontekst, õiguslik raamistus, majandusmõjud, sotsiaalkultuuriline kontekst ja ajakirjanike turvalisus. Eesti pälvis 2021. aastal pressivabaduse edetabelis märkimisväärselt kõrge neljanda koha Norra, Rootsi ja Taani järel (Reporters Sans Frontieres, 2022).

3. Tekstikaeve ja klassifitseerimismudelid

3.1 Tekstikaeve

Andmekaeve (*data mining*) on kiiresti arenev valdkond, sest tarkvaratehnoloogia areng on kaasa toonud andmete hea kättesaadavuse. Veebi vahendusel on võimalik säilitada suurt hulka andmeid (VanderPlas, 2016). Struktureeritud andmed asuvad arvutustabelites või andmebaasides. Tekstid, pildid, heli ja video on struktureerimata andmed, mida on masinate abil keeruline analüüsida struktuuri puudumise tõttu. Poolstruktureeritud andmed saab veebis talletada XML-märgendite (*Extensible Markup Language*) abil (Gandomi & Haider, 2015). Gandomi ja Haideri järgi suurandmete käsitlemine jaguneb andmehalduseks ja analüüsiks. Andmehaldus on esmalt andmete kogumine ja salvestamine, analüüsiks vajalike osade eraldamine, puhastamine, märgendamine ja esitlus. Analüüs sisaldab mudelite loomist ja järelduste tegemist.

Tekstikaeve peamine eesmärk on teabe analüüsimine mustrite avastamiseks tekstist (VanderPlas, 2016). Tekstikujul andmed tekivad sotsiaalvõrgustikest, foorumitest, blogidest, uudistest, e-kirjadest, ametlikest dokumentidest. Tekstianalüüs sisaldab tekstiandmete statistilist analüüsi, keeletehnoloogiat ja masinõpet (Gandomi & Haider, 2015). Keeletehnoloogia tegeleb keelenähtuste modelleerimisega ja loomuliku keele töötusega (*natural language processing*).

Tekstikaeves on sõltuvalt tekstide (dokumentide) uurimise eesmärgist kasutuses mitmeid erinevaid meetodeid (Aggarwal & Zhai, 2012):

Informatsiooni eraldamine tekstidest (*information extraction*) – mittestruktureeritud tekstidest tähenduslike ja mõtestatud seoste tuvastamine, eraldatud informatsioon on struktureeritud.

Tekstist kokkuvõtte tegemine (*text summarization*) – algtekstist tehakse kokkuvõtte või sünteesitakse kokkuvõtteks uus teabeühik. Tekstiliste andmete puhul on vaja tekstikujul olev informatsioon muuta numbriliseks. TF-IDF (*term frequency–inverse document frequency*) meetod annab sõnadele kaalud sõnade sageduste esinemiste järgi.

Juhendamata masinõpe (*unsupervised learning methods*) – meetod ei nõua treeningandmeid ja tekstiandmete märgendamist. Klasterdamine (*clustering*) paigutab tekstid klastritesse tekstides leiduvate seaduspärasuste alusel. Teemade modelleerimine (*topic modeling*) liigitab tekstid tõenäosuste alusel erinevate teemade vahel.

Juhendatud masinõpe (*supervised learning methods*) – meetod vajab märgendatud treeningandmeid. Tekstide klassifitseerimiseks on tarvis treeningandmeid, mille kohta on teada

iga dokumendi klass. Treeningandmeid kasutades konstrueeritakse klassifitseeriv mudel. Saadud mudelit rakendatakse testandmetele ja ennustatakse dokumendi klassi. Klassifitseerimist kasutatakse sageli uudiste filtreerimises ja kategooriatesse jaotamisel, kasutajate hinnangute kogumisel, e-kirjade klassifitseerimisel ja spämm-kirjade filtreerimiseks. E-kirjade klassifitseerimist Maanteeameti näitel kasutas Risto Hinno oma magistritöös (Hinno, 2018).

EstNLTK on eestikeelsete tekstide jaoks loodud Pythoni teekide kogu. EstNLTK abil saab teha tekstide sõnestamist, morfoloogilist analüüsi, lemmatiseerimist, nimeüksuste tuvastamist (Orasmaa, Petmanson, Tkachenko, Laur, & Kaalep, 2016). Käesolevas töös kasutatakse EstNLTK abi nimeüksuste tuvastamisel, tekstide sõnestamisel ja lemmatiseerimisel ehk algvormide leidmisel.

3.2 Tekstide representeerimine

Enamik tekstikaeve meetoditest toetub sõnahulkade põhisele (*bag-of-words*) esitusviisidele, mis kaotab dokumentidest sõnade omavahelise semantilise teabe. Sõnapõhist lähenemist kasutatakse klassifitseerimisülesandes algoritmi lihtsuse tõttu (Aggarwal & Zhai, 2012).

Enamik masinõppe algoritme vajab mudelite loomiseks kindla pikkusega numbrilisi tunnuste vektoreid. Vektoriseerimine on protsess, kus tekstikorpused ehk tekstidokumendid viiakse numbrilisele kujule, loendatakse tekstidokumendis olevad sõnad ja väljastatakse korpuses olevate sõnade sageduste maatriks (scikit-learn, 2022).

Tekstiandmete puhul tuleb arvestada suurte ja hõredate andmemassiividega. Tekstikorpused paigutatakse sõna-dokumentmaatriksisse (*term-document matrix*), kus igale reale vastab üks dokument korpuses ja tunnusteks on dokumentides esinevad sõnad. Sellisel viisil saadakse $d \times n$ mõõtmetega maatriks, kus d on dokumentide arv korpuses ja n on kõigis dokumentides esinevate sõnade arv. Iga dokumendi kohta loendatakse erinevate sõnade esinemise arv, antud dokumendis mitteesineva sõna väärtuseks on 0. Tulemuseks on hõre andmemaatriks, mis sisaldab palju nulle.

Tfidf on kaalud, millega mõõdetakse sõnade olulisust. Sõna t sagedus dokumendis d leitakse valemis 1 (Jurafsky & Martin, 2021):

$$tf_{t,d} = \log(\text{count}(t, d) + 1) \quad (1)$$

Dokumendi pöörsagedus *idf* arvutatakse valemis 2:

$$idf_t = \log\left(\frac{N}{df_t}\right) \quad (2),$$

kus N on dokumentide arv ja df_t on dokumentide arv, kus sõna t esineb.

Kaalud $tfidf$ (valem 3) on eelnevalt leitud väärtuste korrutis

$$tfidf = tf_{t,d} \times idf_t \quad (3)$$

Käesolevas töö kasutatakse $tfidf$ kaale logistilise regressiooni ja tugivektorklassifitseerija mudelis.

Sarnases kontekstis olevatel sõnadel on tavaliselt sarnane tähendus. Sõnade omavahelise semantika ehk tähenduse esiletoomiseks kujutatakse sõna punktina mitmemõõtmelises ruumis. Sõnavektori (*word-embeddings*) moodustamisel arvestatakse naabersõnadega. Sõnavektorite dimensioonidel ei ole selget tõlgendust, sõnavektorite koordinaadid võivad olla ka negatiivsed. Kui sõna-dokumendimaatriks on hõre maatriks paljude nullidega, siis sõnavektorid muudavad vektorid tihedaks, kompaktsemaks ja lühemaks. Tihedad sõnavektorid töötavad teksti analüüsimisel paremini kui hõredad maatriksid. Hõredates andmemaatriksites on näiteks sõna „autobiil“ ja „sõiduauto“ erinevad sõnad, kuid tihedad vektorid leiavad paremini ülesse sünonüümid (Jurafsky & Martin, 2021).

3.3 Klassifitseerimismudelid

3.3.1 Logistiline regressioon

Loomuliku keele töötuses on sagedasti kasutusel klassifitseerimismudelina logistilise regressiooni meetod (*logistic regression*), mille abil saame tekste liigitada binaarselt („positiivne hoiak“ ja „negatiivne hoiak“) või ühte paljudest klassidest. Logistilise regressiooni meetod kasutab eristavat (*discriminative*) raamistikku mudeli loomiseks. Eristav mudel püüab klasse eristama õppida, mitte sisendeid generatiivselt mõista. Generatiivset raamistikku kasutab nt Naïve Bayes klassifitseerimismudel (Jurafsky & Martin, 2021).

Logistilise regressiooni selgitus toetub käesolevas peatükis Jurafsky ja Martini (2021) käsitlusele. Olgu meil üks sisend x , mille tunnused moodustavad vektori (x_1, x_2, \dots, x_n) . Klassifitseerija y väärtus saab olla 1 (sisend kuulub vaadeldavasse klassi) või 0 (sisend ei kuulu vaadeldavasse klassi). Logistiline regressioon toimub kahes faasis. Esmalt treenitakse mudelit, mis tähendab sisendi x tunnustele kaalude w ja väärtuse b leidmist. Seejärel testhulgal arvutatakse tõenäosus $P(y|x)$ ja tagastatakse suurema tõenäosuse saanud märgend $y = 0$ või $y = 1$. Tõenäosus $P(y=1|x)$ tähendab, et otsus on „positiivse hoiakuga“ ja $P(y = 0|x)$ tähendab „negatiivse hoiakuga“ otsust. Treeninghulgal sobitab logistilise regressiooni mudel sisendi x tunnustele kaalude vektori w ja mudeli vabaliikme b . Kaal w_i näitab sisendi vastava tunnuse

olulisust klassifitseerimise otsuse juures. Iga x_i kaalutakse kaaluga w_i , korrutised summeeritakse ning lisatakse vabaliige, saadakse väärtus z :

$$z = \left(\sum_{i=1}^n w_i x_i \right) + b \quad (4)$$

Valem 4 vektorkujul:

$$z = \mathbf{w} \cdot \mathbf{x} + b \quad (5)$$

Logistilise regressiooni mudelis kasutatakse tõenäosuse leidmiseks S-tähte meenutava kujuga sigmoidfunktsiooni ehk logistilist funktsiooni:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (6)$$

Logistilise funktsiooni argumentiks sobib iga reaalarv, funktsiooni $\sigma(z)$ väärtused on lõigus $[0,1]$.

Tõenäosuse $P(y=1)$ leidmiseks saame:

$$P(y = 1) = \sigma(\mathbf{w} \cdot \mathbf{x} + b) = \frac{1}{1 + e^{-(\mathbf{w} \cdot \mathbf{x} + b)}} \quad (7)$$

Tõenäosuse $P(y=0)$ leidmiseks:

$$P(y = 0) = 1 - P(y = 1) = \frac{e^{-(\mathbf{w} \cdot \mathbf{x} + b)}}{1 + e^{-(\mathbf{w} \cdot \mathbf{x} + b)}} \quad (8)$$

Otsuse, millisesse klassi antud sisend x kuulub, saame teha nüüd selliselt:

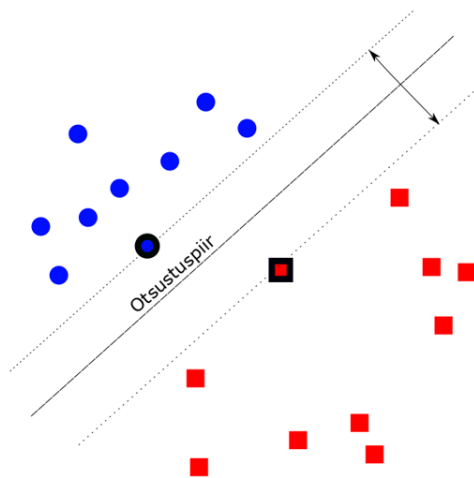
$$otsus(x) = \begin{cases} 1, & \text{kui } P(y = 1|x) > 0,5 \\ 0, & \text{muudel juhtudel} \end{cases} \quad (9)$$

Kui tõenäosus $P(y=1|x)$ on suurem kui 0,5, siis sisend x kuulub klassi $y=1$, vastasel juhul kuulub sisend klassi $y=0$.

3.3.2 Tugivektorklassifitseerija

Tugivektorklassifitseerija (*Support Vector Machine - SVM*) mudel püüab andmete vahele paigutada optimaalseid lineaarseid kui ka mittelineaarseid klassipiire nii, et andmed eristuksid kõige paremini. SVM meetod on sarnaselt logistilise regressiooni mudelile sisendeid eristav (*discriminative*) raamistik (Aggarwal & Zhai, 2012). Joachimsi (1998) järgi sobib SVM mudel tekstiandmete jaoks, sest meetod töötab hästi suurte ja hõredate andmematriksitega. Samuti on tekstide puhul enamasti klassipiirid lineaarsed.

Otsustuspiir (*optimal hyperplanes*) on lineaarne funktsioon, mis on võimalikult maksimaalsel kaugusel kahe klassi elementidest. Optimaalse otsustuspiiri konstrueerimisel on täheldatud, et tuleb arvestada vaid väikese hulga treeningandmetega, mida nimetatakse tugivektoriteks (Cortes & Vapnik, 1995). Tugivektorklassifitseerija otsib sisendeid, mis asuvad klasside äärealadel. Joonisel 1 on sisendid kujutatud kahemõõtmelises ruumis siniste ringidega ja punaste ruutudena. Andmete vahel on otsustuspiir, mis on parima eraldusvõimega antud andmete jaoks. Tugivektoriteks olevad andmepunktid on märgitud tumedama piirjoonega. Tugivektorid defineerivad maksimaalse piiriala (*margin*) ehk klassidevahelise marginaali. Joonisel 1 on maksimaalne piiriala kahe paralleelse punktiirjoone vahel.



Joonis 1. Tugivektorklassifitseerija otsustuspiir ja tugivektorid.

Olgu optimaalseks otsustuspiiriks tasand

$$\mathbf{w}_0 \cdot \mathbf{x} + b_0 = 0 \quad (10),$$

kus \mathbf{w}_0 on otsustuspiiriks oleva tasandi kaalude vektor. Kui märgendatud treeningandmed on

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_d, y_d), \quad \text{kus } y \in \{-1, 1\} \quad (11),$$

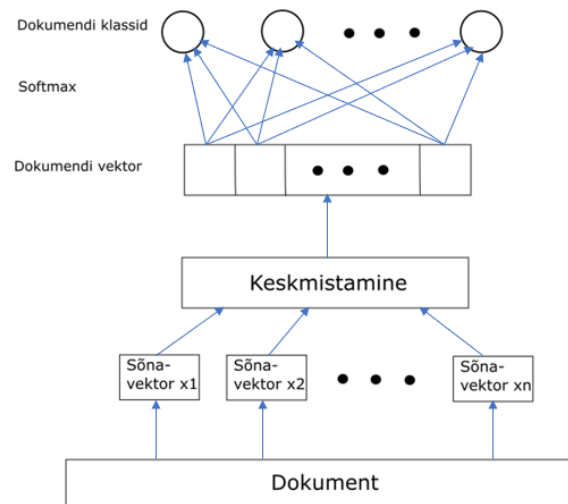
siis maksimaalse piiriala määravad võrratused (Cortes & Vapnik, 1995)

$$\begin{cases} w \cdot x_i + b \geq 1, & \text{kui } y_i = 1 \\ w \cdot x_i + b \leq -1, & \text{kui } y_i = -1 \end{cases} \quad (12)$$

3.3.3 Facebook fastText mudel

Närvivõrkudel põhinevad mudelid on muutunud järjest populaarsemaks, kuid närvivõrkude puuduseks on aeglus mudelite treenimisel ja testimisel (Joulin, Grave, & Bojanowski, 2016). fastText on avatud lähtekoodiga vahend, mille loojaks on Facebook AI Research lab.

fastText on lihtne lineaarne klassifikaator, mis võimaldab treenida mudeleid täpsuseni, mis on võrreldav keerukamate süvaõppemudelitega, kuid mudeli treenimiseks ja testimiseks kuluv aeg on väiksem. Dokumendi sõnadest luuakse sõnavektorid, mis keskmistatakse ja tekib dokumendi vektor. Softmax funktsioon arvutab tõenäosused klassidest, kuhu antud dokument kuulub. Saadud tulemust võrreldakse dokumendi tegeliku klassiga (Zolotov & Kung, 2017).



Joonis 2. fastText klassifitseerimine.

fastText on implementeeritud C++ keeles. Antud töös on kasutatud fastText'i Python teeki.

3.4 Mudelite hindamise meetodid

Segadusmaatriks (*confusion matrix*) väljendab saadud mudeli headust. Maatriksi peadiagonaalil (tabel 1) asetsevad kõik klassifikaatori poolt õigesti ennustatud väärtused: õiged positiivsed (*true positive* – TP), õiged negatiivsed (*true negative* – TN). Peadiagonaalilt välja jäänud on valesti ennustatud väärtused: vale positiivsed (*false positive* – FP), vale negatiivsed (*false negative* – FN).

Tabel 1. Segadusmaatriks.

		Ennustatud märgend	
		pos	neg
Tegelik märgend	pos	TP	FN
	neg	FP	TN

Mudelite hindamise mõõdikud:

$$täpsus = \frac{TP}{TP + FP} \quad (13)$$

$$saagis = \frac{TP}{TP + FN} \quad (14)$$

$$F_1 = \frac{2 \cdot täpsus \cdot saagis}{täpsus + saagis} \quad (15)$$

Täpsus (*precision*) näitab õigete positiivsete väärtuste osakaalu (TP) kõigi ennustatud positiivsete väärtuste hulgas (TP+FP). Saagis (*recall*) näitab õigete positiivsete väärtuste osakaalu (TP) kõigi tegelike positiivsete väärtuste hulgas (TP+FN). Mida lähemal täpsus ja saagis on arvule üks, seda parem on mudeli ennustusvõime. F₁-skoor arvutatakse täpsuse ja saagise harmoonilise keskmisena. Harmooniline keskmine on nullist erinevate arvude pöördarvude aritmeetilise keskmise pöördarv.

4. Pressinõukogu kaebuste kogumine, töötlemine, ülevaade andmetest

4.1 Kaebuste kogumine ja ettevalmistamine analüüsiks

Pressinõukogule esitatud kaebused on avalikult kättesaadavad aadressilt meedialiit.ee/kaebused/. Esimene Pressinõukogule saabus kaebus esitati 06.07.2002, kaebuste andmed täienevad jooksvalt Meedialiidu veebilehel. Kaebuse kohta saab järgmist teavet: kaebaja nimi ja kaebuse esitamise kuupäev, väljaanne (artikli, saate vms pealkiri ja kuupäev), kaebuse sisu, lahendi tulemus ja otsuse sisu. Meedialiidu veebilehelt koguti veebiämbliku abil kaebuste andmestik.

Käesolevas töös on vaatluse all 1081 Pressinõukogule aastatel 2002-2021 esitatud kaebust. Allolevas tabelis 2 on näidiseks toodud kahe kaebuse kohta teadaolevad andmed. Kogutud kaebuste andmestikus on kuus tunnust (veergu):

- Jrk nr;
- Kaebaja – kaebaja nimi, kaebuse kuupäev;
- Väljaanne – väljaande nimi, ilmumise kuupäev, pealkiri;
- Kaebuse sisu;
- Lahend – lahendi tulemus, lahendi kuupäev;
- Otsuse sisu – õigeksmõiste, tauniv või muu lahend koos põhjendusega ja viidetega Eesti ajakirjanduseetika koodeksi punktidele.

Tabel 2. Näide kaebuste tabeli andmeväljadest.

#	Kaebaja	Väljaanne	Kaebuse sisu	Lahend	Otsuse sisu
1081	<i>Kaebaja nimi</i> 27.12.21	Pärnu Postimees 07.12.21 Kare pakane viis bussid rivist välja, autoabi osutajatel käed-jalad tööd täis	Kaebaja leiab, et artikli foto on eksitav, sest jätab mulje nagu poleks <i>kaebaja nimi</i> sõita saanud.	Pressinõukogu otsus 27.01.22	ÕIGEKSMÕISTEV: Pärnu Postimees ei eksinud ajakirjanduseetika koodeksi vastu. Pressinõukogu hinnangul ei jäta artikkel muljet, nagu oleks <i>kaebaja nimi</i> bussidel käreda külmaga probleeme olnud. Ka pildiallkiri ütleb, et liinil nr 40 probleeme ei olnud.
1079	<i>Kaebaja nimi</i> 08.12.21	Hiiu Leht 03.12.21 Maksumaksjale läks Valge Ingli afäär maksma 51 000 eurot	Kaebaja leiab, et artiklis ja juhtkirjas on tema kohta esitatud tõsist kriitikat, kuid sõna ei antud. Kaebaja hinnangul sisaldab artikkel eksitavat ja ebatäpset infot.	Pressinõukogu otsus 27.01.22	TAUNIV: Hiiu Leht rikkus ajakirjanduseetika koodeksi punkti 5.1., mis näeb ette, et kui kellegi kohta avaldatakse tõsiseid süüdistusi, tuleks talle pakkuda kommentaari võimalust samas numbris. Pressinõukogu hinnangul süüdistatakse artiklis <i>kaebaja nimi</i> tuuliku naabrite hirmutamises ja eksitava info jagamises. Samuti rikkus leht ajakirjanduseetika koodeksi punkti 3.3., mis näeb ette, et ajakirjanik peab rangelt kinni informatsiooniallikale antud lubadustest ja väldib lubadusi, mida ta ei suuda täita. Pressinõukogu hinnangul ei pidanud ajakirjanik kinni <i>kaebaja nimi</i> antud lubadusest avaldada tema seisukoht järgmises artiklis.

Kaebuste statistilise analüüsi jaoks on vaja andmestiku tekstiväljadelt eraldada huvipakkuv info. Info eraldamiseks ja kogu edasiseks analüüsiks kasutatakse Jupyter notebook'i keskkonda.

Kaebuste andmestiku tunnuse Kaebaja tekstiväljadelt eraldatakse kaebuse esitamise kuupäev ja kaebuse esitamise aasta ning lisatakse andmestikku kaks tunnust: kaebuse esitamise kuupäev ja kaebuse esitamise aasta.

Kaebuste andmestiku tunnuse Väljaanne tekstiväljadelt eraldatakse väljaande nimi ja määratletakse väljaande liik: kirjutav meedia (trükimeedia) või rääkiv meedia (tele, raadio). Väljaande nimi seotakse väljaande liigiga. Väljaanne, mis esindab nii trükimeediat kui rääkivat meediat, seotakse domineeriva väljaande liigiga (nt ERR – rääkiv meedia, Postimees – trükimeedia). Väljaande nimi eraldatakse kahel viisil:

- Tekstist eraldatakse väljaande nimi etteantud väljaannete nime põhjal.
- Tekstist eraldatakse väljaande nimi nimeüksuste tuvastaja abil.

Kahel viisil väljaande nime eraldamise eesmärgiks on saada võimalikult täpsed andmed väljaande nime ja väljaande liigi osas. Kaebuste andmestikku lisatakse kaks tunnust – väljaande nimi ja väljaande liik.

Kaebuste andmestiku tunnuse Lahend tekstiväljadelt eraldatakse lahendi kuupäev ja lahendi tulemus ning lisatakse kaebuste andmestikku kolm uut tunnust: Pressinõukogu tehtud lahendi kuupäev, lahendi tulemus1 – binaarne tunnus väärtustega „pressinõukogu otsus“ või „otsust ei tehtud“ ja lahendi tulemus2 – Pressinõukogu tehtud lahendi tulemused.

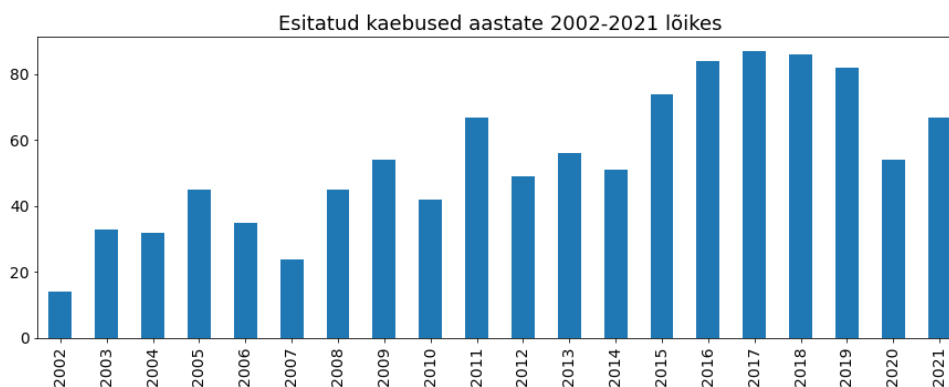
Kaebuste andmestiku tunnuse Otsuse sisu tekstiväljadelt eraldatakse Pressinõukogu otsuses mainitud koodeksipunktid. Andmestikku lisatakse tunnus Koodeks otsuse tekstiväljalt eraldatud koodeksipunktidega. Tunnuse Koodeks abil lisatakse andmestikku kõikidele mainitud koodeksipunktidele vastavad tunnused, kus „1“ tähistab vastava koodeksipunkti leidumist antud otsuses ja „0“ mitteleidumist. Vaadeldaval perioodil on Pressinõukogu otsustes nimetanud 26 erinevat ajakirjanduseetika koodeksipunkti. Kaebuste andmestikku lisatakse 26 uut tunnust.

Kaebuste andmestiku tunnuse Otsuse sisu tekstiväljadelt eraldatakse otsuse liik. Otsuste liik saab olla kas õigeksmõistev, tauniv, muu lahend või topeltlahend. Andmestikku lisatakse iga kaebuse kohta uus tunnus Otsuse liik.

Pärast info eraldamist algsest andmestikust ja uute tunnuste lisamist kasutatakse edasises analüüsis andmestikku, milles on kokku 41 tunnust.

4.2 Statistiline ülevaade kaebustest

Pressinõukogule esitatud kaebuste hulk aastas (joonis 3) oli Pressinõukogu tegutsemise algusaastatel alla 50 kaebuse. Kaebuste hulk oli mõnevõrra suurem aastatel 2016-2019, lähenedes 90 kaebusele aastas. Keskmiselt esitatakse Pressinõukogule 54 kaebust ühes aastas, igas kuus keskmiselt 4-5 kaebust.



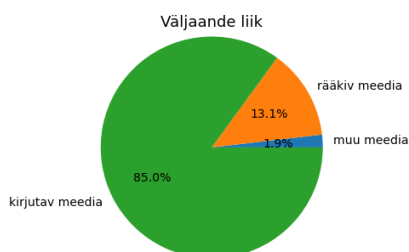
Joonis 3. Pressinõukogule esitatud kaebused aastate 2002-2021 lõikes.

Tabelist 3 selgub, et Pressinõukogu otsuse, kas õigeksmõistva või tauniva, on saanud vaadeldaval perioodil 923 kaebust, mis on 85,4% kõigist esitatud kaebustest. Ligi 15% kõigist Pressinõukogule esitatud kaebustest ei saa Pressinõukogult õigeksmõistvat või taunivat otsust, vaid lahendi tulemuseks nimetatakse erinevas sõnastuses kokkuleppeid, tagasilükkamist, tagasi võtmist, menetluse lõpetamist jms.

Tabel 3. Pressinõukogule esitatud kaebuste lahendite tulemused.

Lahendi tulemus ²	Lahendite arv
pressinõukogu otsus	923
tagasi lükatud	59
tagasi võetud	20
kokkulepe	18
menetlus lõpetatud	13
poolte kokkulepe	13
lahendatud	12
lahendita	10
kokkulepe eelmenetluse käigus	8
jäetud läbi vaatamata	2
pooleid leppisid kokku	1
eelkokkulepe	1
lõpetatud	1
Kokku	1081

Kirjutavas meedias ilmunud artiklite kohta on Pressinõukogule esitatud 919 kaebust, mis on kõigist kaebustest 85,0% (joonis 4), rääkiva meedia saadete kohta on esitatud 142 kaebust (13,1%). Tekstiväljadelt automaatselt väljaande nime ja meedialiigi eraldamine ei õnnestunud 20 kaebuse korral (muu meedia). Valdav osa Pressinõukogule esitatud kaebustest pärinevad kirjutavast meediast.



Joonis 4. Esitatud kaebused väljaande meedialiigi lõikes.

Tabelist 4 on näha, et enim kaebusi on esitatud järgmiste väljaannete kohta: Postimees (15,6%), Eesti Ekspress (9,4%), Eesti Päevaleht (9,4%), Delfi (9%), Õhtuleht (7,6%), Äripäev (5,1%), moodustades kõigist kaebustest 56%. Rääkiva meedia kanalitest on kaebusi esitatud Kanal2, ERR-i, TV3 ja Kuku Raadio kanalites ilmunud saadete kohta. Enamasti esitatakse kaebusi konflikte ja vastuolusid uurivate saadete (Pealtnägija, Kuuuuriija, Radar) ning uudistesaadete kohta (Aktuaalne Kaamera, Reporter, Seitsmesed Uudised). Vähemal määral on kaebusi esitatud piirkondlikes ja maakondlikes ajakirjanduslikes väljaannetes ilmunud artiklite kohta. Täielik nimekiri kõigi esitatud kaebuste kohta väljaannete lõikes asub käesoleva töö Lisas.

Tabel 4. Esitatud kaebused väljaannete lõikes.

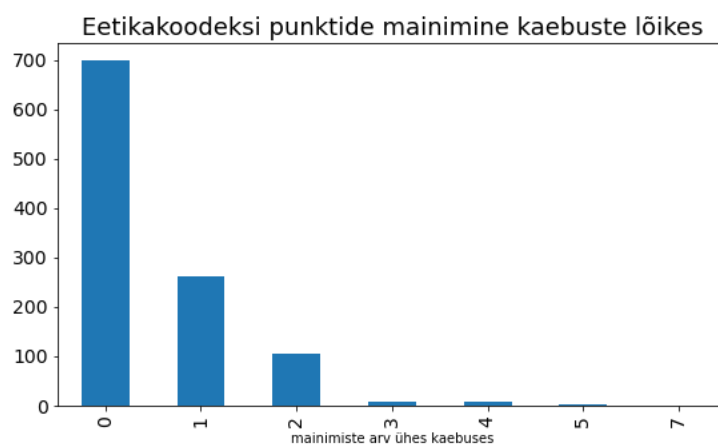
Väljaande nimi	Kaebusi	Osakaal, %
Postimees	169	15,6
Eesti Ekspress	102	9,4
Eesti Päevaleht	102	9,4
Delfi	97	9
Õhtuleht	82	7,6
Äripäev	55	5,1
Kanal2	54	5
ERR	48	4,4
TV3	35	3,2
Pärnu Postimees	23	2,1
Lääne Elu	20	1,9
Saarte Hääl	19	1,8
Meie Maa	16	1,5
väikese osakaaluga	259	24,0
Kokku	1081	100

Eetikakoodeks on kirjalike normide või väärtuste kogumik, mis toetavad teatavas valdkonnas moraaliprobleemide lahendamist ja aitavad leida õigeid käitumisviise (Eetikaveeb, 2007). Pressinõukogu lähtub kaebuste lahendamisel heast ajakirjandustavast, sealhulgas Eesti ajakirjanduseetika koodeksist. Eesti ajakirjanduseetika koodeksis on 36 koodeksipunkti, mis jagunevad kuue valdkonna vahel:

1. Üldised põhimõtted.
2. Sõltumatus.
3. Ajakirjanik ja informatsiooniallikas.
4. Avaldamisreeglid.
5. Vastulause.
6. Reklaam.

Iga valdkonna alla kuulub kolm kuni 12 eetikakoodeksi punkti. Kõige enam koodeksipunkte on avaldamisreeglite valdkonna all. Eesti ajakirjanduseetika koodeks asub käesoleva töö Lisas, koodeks on avalikult kättesaadav veebilehelt meedialiit.ee/eetikakoodeks/.

Pressinõukogu otsustes on mainitud ajakirjanduseetika koodeksipunkte 382 kaebuse korral, koodeksipunktide mainimisi ei ole 699 kaebuse korral (joonis 5). Koodeksipunktide mainimise korral nimetatakse Pressinõukogu otsustes valdavalt 1-2 koodeksipunkti.



Joonis 5. Ajakirjanduseetika koodeksipunktide mainimine kaebuste lõikes.

Suurim ajakirjanduseetika koodeksipunktide mainimine ühes kaebuses on olnud seitse mainimist ühe otsuse kohta. Kahe kaebuse korral on eetikakoodeksi punkte mainitud viis korda ühe kaebuse kohta. Järgnevas tabelis 5 on väljatoodud koodeksipunktide suurimate mainimiste arvuga kaebused.

Tabel 5. Koodeksipunktide suurimate mainimiste arvuga kaebused.

Väljaanne	Kaebuse sisu	Otsuse sisu	Koodeksipunktide mainimisi
Postimees 09.12.14 Abi sotsidele tuleb papa Ossinovski vahendusel Venemaalt	Kaebaja hinnangul on artikli pealkiri eksitav ja halvustav ning selle puhul on tegemist tõsise süüdistusega. Kaebaja ei ole rahul, et artiklis nimetatud erakonna liikmetelt ei ole kommentaari küsitud ja neile on omistatud väide, mida nad ei ole väljendanud.	TAUNIV: Postimees rikkus ajakirjanduseetika koodeksi punkti 1.4., mis näeb ette, et ajakirjandusorganisatsioon kannab hoolt selle eest, et ei ilmuks ebatäpne, moonutatud või eksitav info. Pressinõukogu hinnangul eksitab artikkel lugejaid, jättes mulje, nagu rahastatakse Sotsiaaldemokraatliku Erakonda Oleg Ossinovski kaudu Venemaalt, kuigi erakonda toetab Oleg Ossinovski eraisikuna. Samuti rikkus leht punkte 3.5, 4.1, 4.2, 5.1, 4.3 ja 4.11.	7
Eesti Ekspress 16.09.10 Kris Taska ei taha palka maksta	Kaebaja hinnangul on artikkel moonutatud, ei vasta tegelikkusele ja sisaldab faktivigu. Samuti ei ole kaebaja rahul, et lehe veebis tema seisukohti ei avaldatud.	ÕIGEKSMÕISTEV: Eesti Ekspress ei eksinud antud artiklis ajakirjanduseetika koodeksi punktide 1.5., 4.1., 4.2., 3.5. ja 5.1. vastu, nagu väidab kaebaja. Kaebaja ei osuta konkreetsetele faktivigadele ning on saanud artiklis esitatud etteheidete osas piisaval määral sõna. Pressinõukogu hinnangul on lehel kinnitusi filmitegijatele maksmata arvete ja töötasude kohta, et anda alust sellest kirjutamiseks.	5
Postimees 08.04.08 Vanglajuht Arvi Laike lasi vangidel oma Läänemaa kodu sisustada; lehe kommentaari Valel on lühikesed jalad	Kaebaja hinnangul sisaldab artikkel valefakte, kontrollimata ja kallutatud infot ning kahjustab tema kui vanglajuhi mainet ja usaldusväarsust.	TAUNIV: Pressinõukogu otsustas, et Postimees on artikliga rikkunud ajakirjanduseetika koodeksi punkte 3.5., 4.1., 4.2., 5.1. Toimetuse kommentaari „Valel on lühikesed jalad“ osas otsustas Pressinõukogu, et sellega on rikutud eetikakoodeksi punkti 4.4. Pressinõukogu peab kohatuks, et Postimees on andnud oma kommentaaris hinnangu selle kohta, et Arvi Laike oli julgust Pressinõukogu poole pöörduda.	5

Vaadeldaval perioodil on Pressinõukogu otsustes nimetanud 26 erinevat ajakirjanduseetika koodeksipunkti. Kokku on kaebuste andmestikus 535 korral mainitud ajakirjanduseetika koodeksipunkte. Kõige sagedasem on koodeksipunkti 4.2 mainimine, mis ütleb, et konflikti sisaldava materjali puhul peab ajakirjanik ära kuulama kõik osapooled. Sageli mainitakse Pressinõukogu otsustes koodeksipunkte 1.4, 4.11 ja 4.1. Koodeksipunkti 1.4 järgi vastutab ajakirjanik oma sõnade ja loomingu eest ning ajakirjandusorganisatsioon kannab hoolt, et ei ilmuks ebatäpne, moonutatud või eksitav informatsioon. Fotod, fotode allkirjad, pealkirjad, juhtlauseid ega saatetutvustused ei või auditooriumi eksitada – nimetab koodeksipunkt 4.11. Koodeksipunktis 4.1 on väljatoodud, et uudised, arvamused ja oletused olgu selgelt eristatavad ning uudismaterjal põhinegu tõestataval ja tõenditega tagatud faktilisel informatsioonil.

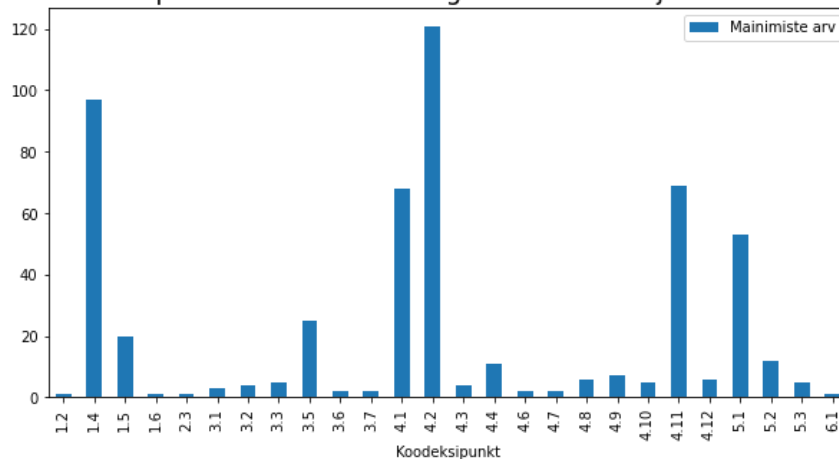
Tabelist 6 on näha, et Pressinõukogult on saanud õigeksmõistva otsuse 488 kaebust (45,1%). Tauniva otsuse (421 kaebust) on saanud 38,9% kaebustest. Muu lahend on leitud 170 kaebuse korral (15,7%). Kaks otsust on sõnastatud kahte moodi mõistetavalt (topeltlahend).

Tabel 6. Otsuse liigid.

Otsuse liik	Otsuste arv
õigeksmõistev	488
tauniv	421
muu otsus	170
topeltlahend	2
Kokku	1081

Järgnevas analüüsis vaadeldakse kaebusi, mis on Pressinõukogult saanud õigeksmõistva või tauniva otsuse (909 kaebust). Nimetame seda andmestikku edaspidi Kaebused_1. Andmestikus Kaebused_1 on 533 korda mainitud 26 erinevat ajakirjanduseetika koodeksipunkti (joonis 6). Valdav enamus ajakirjanduseetika koodeksipunktide mainimisi leiab aset õigeksmõistvate ja taunivate otsustega kaebuste korral. Õigeksmõistvate otsuste korral on koodeksipunkte mainitud 12 korral, taunivate otsuste korral 521 korral. Üle viiendiku koodeksipunktide mainimistest andmestikus Kaebused_1 on Pressinõukogu otsustes koodeksipunkti 4.2 märkimine. Üle 18% eetikakoodeksi punktide mainimistest kuulub punktile 1.4, ligi 13% kõigist mainimistest on punktide 4.1 ja 4.11 märkimine. Kolm sagedasemate mainimistega eetikakoodeksi punkti kuulub avaldamisreeglite valdkonna alla. Eetikakoodeksi punktide mainimine koodeksi punktide kaupa on toodud käesoleva töö Lisas.

Eetikakoodeksi punktide mainimine õigeksmõistvates ja taunivates otsustes



Joonis 6. Ajakirjanduseetika koodeksipunktide mainimine õigeksmõistvates ja taunivates otsustes (andmestik Kaebused_1).

Alltoodud tabelis 7 on esitatud väljaannete kaupa õigeksmõistvate ja taunivate otsuste jaotus. Esimese kümne väljaande seas võib märgata, et õigeksmõistva otsuse saanud kaebuste osakaal on kõrgem Õhtulehel. Taunivate otsuste osakaal on esimese kümne väljaande seas suurim Pärnu Postimehel. Postimehe ja Eesti Ekspressi kohta esitatud kaebuste korral on tehtud otsused üle poolte juhtude õigeksmõistvad, Eesti Päevalehe ja Delfi kohta esitatud kaebused on üle poolte juhtude taunivad.

Rääkiva meedia puhul on märgata, et ERR-i saadete kohta tehtud otsused on ülekaalukalt õigeksmõistvad, Kanal2 ja TV3 saadete kohta tehtud otsused jagunevad võrdselt õigeksmõistvate ja taunivate otsuse vahel. Kõigi väljaannete jaotus õigeksmõistvate ja taunivate otsuste kohta on toodud antud töö Lisas.

Tabel 7. Õigeksmõistvate ja taunivate otsuste jaotus väljaannete lõikes.

Väljaande nimi	tauniv	õigeksmõistev	kokku	tauniv, %	õigeksmõistev, %
Postimees	63	80	143	44,1	55,9
Eesti Ekspress	39	56	95	41,1	58,9
Eesti Päevaleht	47	44	91	51,6	48,4
Delfi	49	39	88	55,7	44,3
Õhtuleht	24	45	69	34,8	65,2
Äripäev	21	29	50	42	58
Kanal2	23	22	45	51,1	48,9
TV3	14	14	28	50	50
ERR	11	17	28	39,3	60,7
Pärnu Postimees	13	7	20	65	35
Saarte Hääli	5	13	18	27,8	72,2
Lääne Elu	8	9	17	47,1	52,9
Meie Maa	9	6	15	60	40
Hiiu Leht	8	5	13	61,5	38,5
Võrumaa Teataja	9	4	13	69,2	30,8
Maaleht	4	7	11	36,4	63,6
Harju Elu	5	5	10	50	50
Sakala	3	7	10	30	70
väikese osakaaluga	66	79	145		
Kokku	421	488	909		

Vaadeldes järgnevas tabelis 8 eetikakoodeksi punktide mainimisi taunivates otsustes suuremate väljaannete kaupa, selgub, et need väljaanded saab jagada kahte gruppi. Eesti Päevalehe, Õhtulehe ja Äripäeva artiklite puhul on kõige enam mainimisi saanud koodeksipunkt 1.4. Koodeksipunkti 1.4 mainimine tähendab, et sellesse gruppi kuuluvate väljaannete puhul on tegemist olnud ebatäpse, moonutatud või eksitava informatsiooniga. Õhtulehe puhul lisandub ka koodeksipunkti 4.11 suur osakaal, mille järgi fotod, fotode allkirjad, pealkirjad, juhtlauseid ega saatetutvustused ei või auditooriumi eksitada. Äripäeva puhul võib märgata, et koodeksipunkti 1.4 mainimine on domineeriv, teiste mainitud punktide, sh punkti 4.2 osakaal on väiksem.

Postimehe, Eesti Ekspressi ja Delfi kohta esitatud kaebustes on kõige enam mainimisi saanud ajakirjanduseetika koodeksipunkt 4.2. Koodeksipunkti 4.2 mainimine tähendab, et antud gruppi kuuluvate väljaannete jaoks on olnud probleemiks konflikti sisaldava materjali käsitlemisel kõikide osapoolte ära kuulamine.

Väiksema osakaaluga väljaannete jaoks eetikakoodeksi punktide jaotust eraldi ei ole välja toodud, sest väikese koguse andmete tõttu võib tulemus olla ebatäpne ja moonutatud.

Tabel 8. Suuremate väljaannete taunivates otsustes mainitud eetikakoodeksi punktid.

Domineerib eetikakoodeksi punkt 1.4	Domineerib eetikakoodeksi punkt 4.2																																																		
Ajakirjanik vastutab oma sõnade ja loomingu eest. Ajakirjandusorganisatsioon kannab hoolt selle eest, et ei ilmuks ebatäpne, moonutatud või eksitav informatsioon.	Konflikti sisaldava materjali puhul peab ajakirjanik ära kuulama kõik osapooled.																																																		
<p>Äripäev: taunivates otsustes rikutud eetikakoodeksi punkte</p> <table border="1"> <thead> <tr> <th>Koodeksipunkt</th> <th>Osakaal</th> </tr> </thead> <tbody> <tr><td>1.4</td><td>36</td></tr> <tr><td>1.5</td><td>5</td></tr> <tr><td>4.1</td><td>5</td></tr> <tr><td>4.2</td><td>9</td></tr> <tr><td>4.4</td><td>5</td></tr> <tr><td>4.10</td><td>5</td></tr> <tr><td>4.11</td><td>14</td></tr> <tr><td>5.1</td><td>14</td></tr> <tr><td>5.2</td><td>9</td></tr> </tbody> </table>	Koodeksipunkt	Osakaal	1.4	36	1.5	5	4.1	5	4.2	9	4.4	5	4.10	5	4.11	14	5.1	14	5.2	9	<p>Eesti Ekspress: taunivates otsustes rikutud eetikakoodeksi punkte</p> <table border="1"> <thead> <tr> <th>Koodeksipunkt</th> <th>Osakaal</th> </tr> </thead> <tbody> <tr><td>1.4</td><td>12</td></tr> <tr><td>1.5</td><td>2</td></tr> <tr><td>3.1</td><td>2</td></tr> <tr><td>3.3</td><td>2</td></tr> <tr><td>3.5</td><td>2</td></tr> <tr><td>4.1</td><td>10</td></tr> <tr><td>4.2</td><td>30</td></tr> <tr><td>4.4</td><td>4</td></tr> <tr><td>4.10</td><td>2</td></tr> <tr><td>4.11</td><td>16</td></tr> <tr><td>5.1</td><td>10</td></tr> <tr><td>5.2</td><td>6</td></tr> <tr><td>5.3</td><td>2</td></tr> </tbody> </table>	Koodeksipunkt	Osakaal	1.4	12	1.5	2	3.1	2	3.3	2	3.5	2	4.1	10	4.2	30	4.4	4	4.10	2	4.11	16	5.1	10	5.2	6	5.3	2		
Koodeksipunkt	Osakaal																																																		
1.4	36																																																		
1.5	5																																																		
4.1	5																																																		
4.2	9																																																		
4.4	5																																																		
4.10	5																																																		
4.11	14																																																		
5.1	14																																																		
5.2	9																																																		
Koodeksipunkt	Osakaal																																																		
1.4	12																																																		
1.5	2																																																		
3.1	2																																																		
3.3	2																																																		
3.5	2																																																		
4.1	10																																																		
4.2	30																																																		
4.4	4																																																		
4.10	2																																																		
4.11	16																																																		
5.1	10																																																		
5.2	6																																																		
5.3	2																																																		
<p>Eesti Päevaleht: taunivates otsustes rikutud eetikakoodeksi punkte</p> <table border="1"> <thead> <tr> <th>Koodeksipunkt</th> <th>Osakaal</th> </tr> </thead> <tbody> <tr><td>1.4</td><td>29</td></tr> <tr><td>1.5</td><td>4</td></tr> <tr><td>3.5</td><td>4</td></tr> <tr><td>4.1</td><td>15</td></tr> <tr><td>4.2</td><td>20</td></tr> <tr><td>4.11</td><td>6</td></tr> <tr><td>4.12</td><td>6</td></tr> <tr><td>5.1</td><td>15</td></tr> <tr><td>5.2</td><td>2</td></tr> <tr><td>5.3</td><td>2</td></tr> </tbody> </table>	Koodeksipunkt	Osakaal	1.4	29	1.5	4	3.5	4	4.1	15	4.2	20	4.11	6	4.12	6	5.1	15	5.2	2	5.3	2	<p>Delfi: taunivates otsustes rikutud eetikakoodeksi punkte</p> <table border="1"> <thead> <tr> <th>Koodeksipunkt</th> <th>Osakaal</th> </tr> </thead> <tbody> <tr><td>1.4</td><td>18</td></tr> <tr><td>1.5</td><td>5</td></tr> <tr><td>3.3</td><td>2</td></tr> <tr><td>3.5</td><td>7</td></tr> <tr><td>3.7</td><td>2</td></tr> <tr><td>4.1</td><td>10</td></tr> <tr><td>4.2</td><td>26</td></tr> <tr><td>4.3</td><td>2</td></tr> <tr><td>4.6</td><td>2</td></tr> <tr><td>4.10</td><td>2</td></tr> <tr><td>4.11</td><td>20</td></tr> <tr><td>5.1</td><td>5</td></tr> <tr><td>5.3</td><td>4</td></tr> </tbody> </table>	Koodeksipunkt	Osakaal	1.4	18	1.5	5	3.3	2	3.5	7	3.7	2	4.1	10	4.2	26	4.3	2	4.6	2	4.10	2	4.11	20	5.1	5	5.3	4
Koodeksipunkt	Osakaal																																																		
1.4	29																																																		
1.5	4																																																		
3.5	4																																																		
4.1	15																																																		
4.2	20																																																		
4.11	6																																																		
4.12	6																																																		
5.1	15																																																		
5.2	2																																																		
5.3	2																																																		
Koodeksipunkt	Osakaal																																																		
1.4	18																																																		
1.5	5																																																		
3.3	2																																																		
3.5	7																																																		
3.7	2																																																		
4.1	10																																																		
4.2	26																																																		
4.3	2																																																		
4.6	2																																																		
4.10	2																																																		
4.11	20																																																		
5.1	5																																																		
5.3	4																																																		
<p>Õhtuleht: taunivates otsustes rikutud eetikakoodeksi punkte</p> <table border="1"> <thead> <tr> <th>Koodeksipunkt</th> <th>Osakaal</th> </tr> </thead> <tbody> <tr><td>1.4</td><td>21</td></tr> <tr><td>3.3</td><td>3</td></tr> <tr><td>3.5</td><td>9</td></tr> <tr><td>4.1</td><td>9</td></tr> <tr><td>4.2</td><td>18</td></tr> <tr><td>4.4</td><td>3</td></tr> <tr><td>4.11</td><td>21</td></tr> <tr><td>5.1</td><td>9</td></tr> <tr><td>5.2</td><td>3</td></tr> <tr><td>5.3</td><td>3</td></tr> </tbody> </table>	Koodeksipunkt	Osakaal	1.4	21	3.3	3	3.5	9	4.1	9	4.2	18	4.4	3	4.11	21	5.1	9	5.2	3	5.3	3	<p>Postimees: taunivates otsustes rikutud eetikakoodeksi punkte</p> <table border="1"> <thead> <tr> <th>Koodeksipunkt</th> <th>Osakaal</th> </tr> </thead> <tbody> <tr><td>1.4</td><td>16</td></tr> <tr><td>1.5</td><td>2</td></tr> <tr><td>3.5</td><td>5</td></tr> <tr><td>4.1</td><td>18</td></tr> <tr><td>4.2</td><td>22</td></tr> <tr><td>4.3</td><td>2</td></tr> <tr><td>4.4</td><td>4</td></tr> <tr><td>4.8</td><td>1</td></tr> <tr><td>4.9</td><td>1</td></tr> <tr><td>4.11</td><td>15</td></tr> <tr><td>4.12</td><td>2</td></tr> <tr><td>5.1</td><td>10</td></tr> <tr><td>5.2</td><td>1</td></tr> </tbody> </table>	Koodeksipunkt	Osakaal	1.4	16	1.5	2	3.5	5	4.1	18	4.2	22	4.3	2	4.4	4	4.8	1	4.9	1	4.11	15	4.12	2	5.1	10	5.2	1
Koodeksipunkt	Osakaal																																																		
1.4	21																																																		
3.3	3																																																		
3.5	9																																																		
4.1	9																																																		
4.2	18																																																		
4.4	3																																																		
4.11	21																																																		
5.1	9																																																		
5.2	3																																																		
5.3	3																																																		
Koodeksipunkt	Osakaal																																																		
1.4	16																																																		
1.5	2																																																		
3.5	5																																																		
4.1	18																																																		
4.2	22																																																		
4.3	2																																																		
4.4	4																																																		
4.8	1																																																		
4.9	1																																																		
4.11	15																																																		
4.12	2																																																		
5.1	10																																																		
5.2	1																																																		

4.3 Kaebuste linkide kogumine, tekstikogu loomine

Käesoleva töö oluliseks osaks oli kaebustele vastava tekstikogu loomine, mis oli ajamahukas tegevus. Andmestikust Kaebused_1 tehti automaatne Google-otsing, kasutades tunnust

Väljaanne otsingu märksõnadena. Õigeksmõistva ja tauniva otsuse saanud kaebuste lingid koguti eraldi. Mõlemasse andmestikku lisati uus tunnus Lingid. Seejärel andmestikud ühendati.

Tabelisse 9 on koondatud info tekstidele või saadetele viitavate linkide kohta. Kogutud lingid (909 kaebuse linki) on käsitsi üle kontrollitud. Kontrollimise vajadus selgus leitud linkide asjakohasust üle vaadates. Osutus, et teatud hulk linke viitab valele artiklile. Valed lingid asendati asjakohaste linkidega. Kaebusele viitavat linki ei õnnestunud tuvastada 67 kaebuse korral. Telesaadetele, raadiosaadetele, venekeelsetele artiklitele, fotodele ja reklaamile viitavad lingid eemaldati edasisest analüüsist (163 kaebust). Mõnedes kaebustes ei ole väljatoodud konkreetset artiklit või on nimetatud artikli teema ilma pealkirjata, selliseid kaebusi oli 42.

Tabel 9. Kaebustele viitavad lingid.

	Kaebuste arv
link leitud	614
DIGAR	22
link puudub	67
raadiosaade	6
telesaade	96
venekeelne	56
fotod	4
reklaam	1
tekst puudub	1
teadmata sisu	42
Kokku	909

Leitud linkidega (614) kaebuste seas on esindatud neli suuremat meediakontserni:

- Ekspress Meedia – väljaanded Delfi, Eesti Ekspress, Eesti Päevaleht, Pere ja Kodu, Maaleht, Kroonika – 273 kaebust;
- Postimees Grupp – väljaanded Postimees, Pärnu Postimees, Sakala, Virumaa Teataja, Järva Teataja, Lõuna-Eesti Postimees, Põhjarannik, Saarte Hääl, Oma Saar, Elu24 – 190 kaebust;
- Õhtuleht – 56 kaebust;
- Äripäev – 30 kaebust;
- muud väljaanded – 65 kaebust.

Meediatekstide kogumiseks kasutati veebiämblikku. Eelnevalt nimetatud meediakontsernide tekstid koguti eraldi, sest *html*-vormingud on erinevate väljaannete puhul erinevad. Hiljem andmed ühendati. Artiklite täistekstide saamiseks kasutati iga väljaande jaoks autentimisküpsiseid. Lisaks artikli tekstile koguti artikli pealkiri. Ilmnes, et artiklite pealkirju on pärast kaebuse esitamist muudetud või on kaebuste esitamisel esitatud ebatäpne pealkiri.

Täiendavalt otsiti artiklite tekste DIGAR Eesti artiklite hulgast, nii leiti juurde 22 kaebustega seotud artiklit.

Koguti andmestik, milles on kolm tunnust: kaebuse number, artikli pealkiri ja artikli tekst. Kogutud tekstid ja pealkirjad ühendati andmestikuga Kaebused_1. Ühendatud andmestikus on 636 õigeksmõistvat või taunivat kaebust koos vastava meediatekstiga.

Kogutud meediatekstid puhastati üleliigsetest sümbolitest, *html*-vormingutest, tühikutest. Eemaldati kirjavahemärgid ja muudeti tähed väiketähtedeks. Tekstid sõnestati ja lemmatiseeriti EstNLTK vahenditega järgneva klassifitseerimise jaoks. `fastText` meetodiga klassifitseerimiseks loodi uus tekstifail, kus iga artikli tekst on vastavalt märgendatud `__label__õigeksmõistev` või `__label__tauniv`.

5. Tekstide klassifitseerimine ja analüüs

5.1 Eksperimendi kirjeldus

Kaebustele vastavate meediatekstide klassifitseerimiseks ja mudelite loomiseks kasutati kolme klassifitseerivat meetodit: logistilist regressiooni, tugivektorklassifitseerijat ja fastText klassifitseerimist. Käesolevas töös rakendatud mudelite kirjeldus on toodud alapeatükis 3.3. Tekstide töötlemiseks vajalikud meetodid on kirjeldatud alapeatükis 3.2. Mudelite hindamise mõõdikutest on antud ülevaade alapeatükis 3.4. Kõiki rakendatavaid klassifitseerijaid treeniti nii lemmatiseerimata tekstidel kui ka lemmatiseeritud tekstidel.

Logistiline regressioon ja tugivektorklassifitseerija

Klassifitseerimiseks jaotati uuritavad meediatekstid testandmestikuks (20%) ja treeningandmestikuks (80%). Treeningosal viidi läbi ristvalideerimine (*cross-validation*) optimaalsete hüperparameetrite leidmiseks. Treeningosa ristvalideerimisel modifitseeriti hüperparameetritena järgmisi suurusid: minimaalne sõnasagedus (*min_dff*), maksimaalne sõnasagedus (*max_dff*), sõnamitmikud (*ngram_range*), stoppsõnade list, regulariseeriv hüperparameeter (*C*). Minimaalse sõnasageduse määramisega saab analüüsist kõrvale jätta tekstides harva esinevaid sõnu. Maksimaalse sõnasageduse abil saab analüüsist ignoreerida sõnu, mis on iseloomulikud antud korpusele ja mille esinemine on etteantud lävest kõrgem. Minimaalset ja maksimaalset sõnasagedust saab määratleda nii absoluutarvuna kui osakaaluna. Koosesinevate sõnade ehk sõnamitmike ehk *n*-grammide jaoks saab määrata vahemike pikkused: vahemik (1, 1) tähendab unigramme, vahemik (1, 2) unigramme ja bigramme, vahemik (2, 2) ainult bigramme. Stoppsõnade listiga saab määratleda sõnad, mida soovitakse analüüsist kõrvale jätta, stoppsõnadeks on sageli asesõnad, sidesõnad ja muud sellised sõnad, mis otseselt teksti sisu edasi ei anna. Regulariseeriv hüperparameeter *C* võimendab või pehmenab treeningandmete väärtusi, suur *C* väärtus annab treeningandmetele suurema kaalu. Tugivektorklassifitseerijat treeniti lineaarsel tuumal (*kernel*).

Antud töös modifitseeriti mudelite hüperparameetreid optimaalse parameetrite kombinatsiooni saamiseks järgmiselt:

- minimaalse sõnasageduse väärtusteks {2; 3; 4; 5};
- maksimaalse sõnasageduse osakaaludeks {0,65; 0,75; 0,85; 0,95};
- sõnamitmikud (1, 2) – unigrammid ja bigrammid;
- stoppsõnade list – sidesõnad ja asesõnad;
- regulariseeriv parameeter – {0,1; 1; 2; 5; 6; 10; 20; 50; 100; 500}.

Mudelid treeniti 10-kordse ristvalideerimisega. 10-kordsel ristvalideerimisel jagatakse treeningandmestik kümneks võrdseks osaks, üks osa valitakse ühel korral valideerimisandmeteks, teised osad on treenimisandmeteks. Mudeli sobivuse hindamiseks olid hindamismõõdikuteks tauniva märgendi F_1 -skoor ja kaalutud keskmise F_1 -skoor. Pärast mudeli 10-kordset ristvalideerimist vaadeldavad F_1 -skoorid keskmistati ja fikseeriti mudeli optimaalsed hüperparameetrid. Parimate leitud hüperparameetritega treeniti mudel tervel treeningandmestikul. Testandmestiku põhjal hinnati treenitud mudeli headust tauniva märgendi ja kaalutud keskmise F_1 -skooride põhjal.

fastText

Klassifitseerimiseks jaotati uuritavad meediatekstid testandmestikuks (20%) ja treeningandmestikuks (80%) kahe eelmise mudeliga samadel alustel. fastText mudeli rakendamisel ei toimu ristvalideerimist. Mudeleid treenitakse tervel treeningandmestikul, hüperparameetritest modifitseeriti järgmisi suurusid: epoch (*epoch*), õppimiskordaja ehk algoritmi õppimiskiirus (*learning rate - lr*), sõnavektori pikkus (*dim*) ja eeltreenitud vektorid (*pretrainedVectors*). Epoch on ajavahemik, mille vältel treeningandmed läbivad närvivõrku. Õppimiskordaja näitab mudeli õppimisvõime muutumise kiirust. Eeltreenitud sõnavektorid on treenitud eestikeelse Wikipedia artiklite peal. Käesolevas töös rakendati CBOW-algoritmi (*Continuous Bag of Words Model*) abil eeltreenitud sõnavektoreid ja skip-gram-algoritmi abil eeltreenitud sõnavektoreid (fastText, 2022). Skip-gram algoritm püüab ennustada kontekstisõnu sõna ümber, seevastu CBOW-algoritm ennustab põhisõnu kontekstisõnade abil. Testandmestiku põhjal hinnati treenitud mudeli headust tauniva märgendi ja kaalutud keskmise F_1 -skooride põhjal.

Antud töös modifitseeriti mudelite hüperparameetreid optimaalse parameetrite kombinatsiooni saamiseks järgmiselt:

- epoch {10; 25};
- õppimiskordaja {0,5; 1}
- sõnavektori pikkus {100; 300}
- eeltreenitud vektorid;
- täiendava parameetrina stoppsõnade list – sidesõnad ja asesõnad.

5.2 Tulemused

Logistiline regressioon ja tugivektorklassifitseerija

Ristvalideerimisel saadud F_1 -skoorid koos optimaalsete hüperparameetritega on toodud järgnevas tabelis 10. Logistilise regressiooni mudelite ristvalideerimisel osutus, et lemmatiseerimata tekstide puhul saadi tauniva märgendi F_1 -skooriks parem tulemus kui lemmatiseeritud tekstide puhul. Logistilise regressiooni mudeli kaalutud keskmine F_1 -skoor on kõrgem lemmatiseeritud tekstide korral. Nii lemmatiseeritud kui lemmatiseerimata tekstide puhul olid ühisteks optimaalseteks hüperparameetriteks minimaalne sõnasagedus väärtusega 3, stoppsõnadeks sidesõnad ja asesõnad, sõnamitmikest unigrammid ja bigrammid. Erinev oli regulariseeriv parameeter: lemmatiseerimata tekstide puhul oli optimaalseks regulariseeriva parameetri väärtuseks 100, lemmatiseeritud tekstide korral oli optimaalse regulariseeriva parameetri väärtus 500. Maksimaalse sõnasageduse modifitseerimine mudelites halvendas tulemuste F_1 -skooore nii lemmatiseeritud kui lemmatiseerimata tekstide puhul ja seetõttu jäeti vaatluse alt välja.

Tugivektorklassifitseerija mudelite treenimisel ristvalideerimisega (tabel 10) selgus, et lemmatiseerimata tekstide puhul on tauniva märgendi F_1 -skoor kõrgem lemmatiseeritud tekstide tauniva märgendi F_1 -skoorist. Mudeli kaalutud keskmine F_1 -skoor on samuti kõrgem lemmatiseerimata tekstide puhul. Lemmatiseerimata ja lemmatiseeritud tekstide puhul olid ühisteks optimaalseteks hüperparameetriteks minimaalne sõnasagedus väärtusega 4 ja regulariseeriv parameeter väärtusega 1. Lemmatiseerimata tekstide puhul täienes mudel järgmiste hüperparameetritega: stoppsõnad (sidesõnad ja asesõnad) ja sõnamitmikest unigrammid ja bigrammid. Maksimaalse sõnasageduse modifitseerimine tugivektorklassifitseerija mudelites ei parandanud tulemuste F_1 -skooore.

Logistilise regressiooni mudelite ja tugivektorklassifitseerija mudelite ristvalideerimisel saadud F_1 -skooore võrreldes osutus, et tauniva märgendi F_1 -skoor on kõige kõrgem tugivektorklassifitseerija mudelis lemmatiseerimata tekstide puhul. Logistilise regressiooni mudelites oli optimaalseks minimaalse sõnasageduse väärtuseks 3, tugivektorklassifitseerija mudelites oli optimaalsem minimaalse sõnasageduse väärtus 4. Mõlema mudeli jaoks oli ühine, et maksimaalse sõnasageduse modifitseerimine ristvalideerimisel ei parandanud mudeli omadusi ja jäi seetõttu kõigist mudelitest välja. Ristvalideerimisel saadud mudelite vastavaid F_1 -skooore võrreldes on näha, et leitud skooride erinevused ei ole suured.

Tabel 10. Ristvalideerimisel saadud F₁-skoorid.

mudel	hinnav näitaja	lemmatiseerimata tekst		lemmatiseeritud tekst	
		F ₁ -skoorid	ristvalideerimisel saadud optimaalsed hüperparameetrid	F ₁ -skoorid	ristvalideerimisel saadud optimaalsed hüperparameetrid
Logistiline regressioon	tauniv	0,551	min sõnasagedus 3	0,527	min sõnasagedus 3
	õigeksmõistev	0,694	stoppsõnad	0,670	stoppsõnad
	kaalutud keskmine	0,630	uni- ja bigrammid C = 100	0,655	uni- ja bigrammid C = 500
Tugivektor-klassifitseerija	tauniv	0,553	min sõnasagedus 4	0,542	min sõnasagedus 4
	õigeksmõistev	0,716	stoppsõnad	0,698	
	kaalutud keskmine	0,642	uni- ja bigrammid C = 1	0,627	C = 1

fastText

fastText mudeli korral oli lemmatiseerimata tekstide jaoks parimad hüperparameetrid järgmised: epohh 25, õppimiskordaja 0,5, sõnavektori pikkus 300 ja skip-gram-algoritmi abil eeltreenitud sõnavektor. Lemmatiseeritud tekstide jaoks olid optimaalsed hüperparameetrid järgmised: epohh 10, õppimiskordaja 1 ja sõnavektori pikkus 300, eeltreenitud sõnavektorite kasutamine mudeli hindamismõõdikuid oluliselt ei parandanud. Stoppsõnade eemaldamine tekstidest ei parandanud hindamismõõdikuid lemmatiseerimata ega lemmatiseeritud tekstide jaoks.

Kokkuvõtte mudelite treenimisest

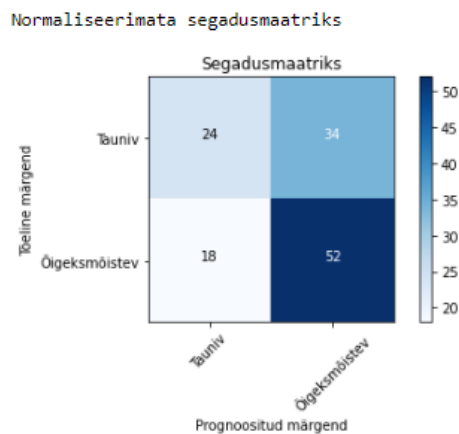
Järgnevas tabelis 11 on kokkuvõtte kõigi rakendatud mudelite F₁-skooridest testandmestikul. Kolme klassifitseerija võrdlemisel tuleb arvestada, et fastText klassifitseerija jaoks on optimaalsed hüperparameetrid valitud testandmestiku tulemuste põhjal ning võivad olla ülehinnatud. Logistilise regressiooni ja tugivektorklassifitseerija mudelites on optimaalsed parameetrid valitud ristvalideerimise teel. Parimate hüperparameetrite abil leitud mudelite F₁-skoorid on suurusjärkude poolest sarnased. Mudelite lõikes prognoosis taunivaid otsuseid kõige paremini fastText mudel lemmatiseeritud tekstidel, halvemini prognoosis taunivaid otsuseid tugivektorklassifitseerija lemmatiseeritud tekstidel. Õigeksmõistvaid otsuseid prognoosis kõige paremini tugivektorklassifitseerija lemmatiseerimata tekstidel ja halvemini fastText mudel lemmatiseerimata tekstidel. Võrreldes lemmatiseerimata ja lemmatiseeritud tekstide hindamismõõdikuid, siis võib märgata, et logistilise regressiooni ja tugivektorklassifitseerija puhul saadi lemmatiseerimata tekstide puhul pisut parem tulemus. fastText mudeli puhul saadi mõnevõrra parem tulemus lemmatiseeritud tekstidega.

Mudeli kaalutud keskmise F_1 -skoori mudeli hindamismõõdikuks aluseks võttes prognoosis vaadeldud mudelitest kõige täpsemini tugivektorklassifitseerija lemmatiseerimata tekstidel.

Tabel 11. Testandmestikul saadud optimaalsete hüperparameetritega F_1 -skoorid.

model	hindamismõõdikud	lemmatiseerimata tekstid	lemmatiseeritud tekstid
Logistiline regressioon	tauniv F_1 -skoor	0,463	0,458
	õigeksmõistev F_1 -skoor	0,683	0,675
	kaalutud keskmine F_1 -skoor	0,584	0,577
Tugivektorklassifitseerija	tauniv F_1 -skoor	0,467	0,452
	õigeksmõistev F_1 -skoor	0,711	0,687
	kaalutud keskmine F_1 -skoor	0,600	0,580
fastText	tauniv F_1 -skoor	0,475	0,480
	õigeksmõistev F_1 -skoor	0,658	0,667
	kaalutud keskmine F_1 -skoor	0,575	0,582

Taunivaid otsuseid prognoosis kõige paremini fastText klassifitseerija lemmatiseeritud tekstidel hüperparameetritega: epohh 10, õppimiskordaja 1 ja sõnavektori pikkusega 300. Joonisel 7 on taunivaid otsuseid kõige paremini prognoosinud mudeli segadusmaatriks. Antud mudeli tauniva märgendiga otsuste prognoosimise täpsus on 0,57 ja saagis 0,41. Mudeli täpsus (valem 13) näitab õigesti prognoositud taunivate otsuste osakaalu kõigi ennustatud taunivate väärtuste hulgas. Mudeli saagis (valem 14) näitab õigesti prognoositud taunivate otsuste osakaalu kõigi tegelike taunivate väärtuste hulgas. Õigesti on prognoositud 24 tauniva otsuse saanud teksti, 34 teksti korral mudel eksis tauniva otsuse prognoosimisel.



Joonis 7. Segadusmaatriks (fastText, lemmatiseeritud tekstid).

5.3 Diskussioon, järeldused

Käesoleva töö esimeseks uurimisküsimuseks oli taunivate otsuste eetikakoodeksi punktide mainimine Pressinõukogu lahendites. Uurimisküsimust käsitleti antud töö peatükis 4. Kõige sagedamini mainiti Pressinõukogu otsustes eetikakoodeksi punkti 4.2, mille järgi konflikti

sisaldava materjali puhul peab ajakirjanik ära kuulama kõik osapooled. Teisel kohal mainimiste arvu poolest oli eetikakoodeksi punkt 1.4, mis ütleb, et ajakirjanik vastutab oma sõnade ja loomingu eest; ajakirjandusorganisatsioon kannab hoolt selle eest, et ei ilmuks ebatäpne, moonutatud või eksitav informatsioon. Sagedamini mainiti ka eetikakoodeksi punkti 4.11 (Fotod, fotode allkirjad, pealkirjad, juhtlauseid ega saatetutvustused ei või auditooriumi eksitada) ja punkti 4.1 (Uudised, arvamused ja oletused olgu selgelt eristatavad. Uudismaterjal põhinegu tõestataval ja tõenditega tagatud faktilisel informatsioonil).

Eetikakoodeksi punktide mainimisi analüüsidest selgus, et suuremate väljaannete lõikes on erinevused eetikakoodeksi punktide mainimiste osas. Ajakirjanduseetika koodeksipunkt 4.2 on saanud enim mainimisi Postimehe, Eesti Ekspressi ja Delfi artiklite kohta esitatud kaebustes. Eesti Päevalehe, Õhtulehe ja Äripäeva artiklite puhul on kõige enam mainimisi saanud koodeksipunkt 1.4.

Konflikti kõigi osapooltele mitte sõnaandmine võib olla seotud publitseeritavate tekstide väljaandmise kiirusega. Väljaanded soovivad olla esimeste seas, kes uudisest kirjutavad ning kõigi osapoolte küsitlemine võib olla seetõttu ajaliselt piiratud. Ebatäpse, moonutatud või eksitava informatsiooni edastamine võib olla samuti seotud ajakirjandusväljaande kiirustamisega meediateksti publitseerimisel. Fakti ja arvamuse eristamine on muutunud järjest keerulisemaks. Meedias on info üleküllus, arvamusi võidakse käsitleda faktidena, mille kontrollimist ei pea ajakirjanik vajalikuks või on fakti mittekontrollimise põhjuseks samuti ajapuudus.

Antud töö teiseks uurimisküsimuseks oli tauniva ja õigeksmõistva otsuse saanud meediatekstide sõnapõhine eristumine. Meediatekste klassifitseerides (alapeatükid 5.1 ja 5.2) selgus, et mõnevõrra paremini eristuvad õigeksmõistvate otsustega tekstid. Taunivate otsustega meediatekstide tuvastamine käesolevas töös leitud mudelitega on vähem edukas. Tekstide klassifitseerimise hindamismõõdikute (F_1 -skooride) põhjal saab järeldada, et Pressinõukogule esitatud kaebustele vastavate meediatekstide klassifitseerimise mudelid ei suuda väga hästi ennustada taunivaid otsuseid.

Põhjusi, miks leitud mudelid tehtud otsuseid väga hästi ei ennusta, võib olla mitmeid. Tehniliselt poolelt saab nimetada, et analüüs toimus väikese andmestikuga (636 artiklit), mis ei pruugi olla treenimisülesandeks piisav suurusega ning mudelite treenimiseks võivad paremini sobida võimsamad eeltreenitud närivõrgupõhised mudelid. Meediatekstide sõnapõhise klassifitseerimise ülesandes ei võeta arvesse artikli laiemat konteksti ega lisainfot,

mis on teada artikli lugejatele ja õigeksmõistvaid või taunivaid otsuseid andvale Pressinõukogule.

Kõige sagedasem eetikakoodeksi punkti mainimine on seotud sellega, et ajakirjanik pole ära kuulnud kõiki osapooli, st konflikti kõik osapooled pole saanud artiklis esitada omi seisukohti ning klassifitseerimisülesanne, mis on oma olemuselt sõnapõhine, võib seetõttu muutuda komplitseerituks ja halvasti ennustavaks.

Veel üheks põhjuseks, miks leitud mudelid ei suuda hästi prognoosida taunivaid otsuseid, võib olla seotud mõnede ajakirjandusväljaannete praktikaga muuta ja parandada juba ilmunud artikleid pärast tauniva otsuse saamist. Selle probleemi lahenduseks oleks kasutada analüüsiks digiteeritud paberväljaanded. Paraku paberväljaannete tähtsus meedias järjest väheneb. Juba ilmunud artiklites faktide parandamine, täiendavate kommentaaride lisamine on lugeja vaates igati positiivne ja näitab, et ajakirjandusväljaanded suhtuvad Pressinõukogu töösse tõsiselt.

6. Kokkuvõte

Antud magistritöö eesmärgiks oli anda statistiline ülevaade Pressinõukogule esitatud kaebustest. Teiseks eesmärgiks oli kaebustele vastavatele meediatekstide põhjal treenida masinõppemeetoditega klassifitseerija, mis eristaks tauniva ja õigeksmõistva otsuse saanud kaebuse. Töö oluliseks osaks oli kaebustele vastavate meediatekstide kogu koostamine ja kogust ülevaate andmine.

Pressinõukogule esitatud kaebusi (2001-2021) analüüsid selgus, et kõige sagedamini mainiti Pressinõukogu otsustes eetikakoodeksi punkti 4.2, mille järgi konfliktis sisaldava materjali puhul peab ajakirjanik ära kuulama kõik osapooled. Eetikakoodeksi punkt 4.2 on saanud enim mainimisi Postimehe, Eesti Ekspressi ja Delfi kohta esitatud kaebustes. Eesti Päevalehe, Õhtulehe ja Äripäeva artiklite puhul on kõige enam mainimisi saanud koodeksipunkt 1.4, mille järgi ajakirjanik vastutab, et ei ilmuks ebatäpne, moonutatud või eksitav informatsioon.

Töö teoreetilises osas anti ülevaade tekstide representeerimisest ja töös rakendatavatest klassifitseerimismeetoditest – logistiline regressioon, tugivektorklassifitseerija ja fastText klassifitseerimine. Mudelid treeniti lemmatiseeritud ja lemmatiseerimata meediatekstidel. Logistilise regressiooni ja tugivektorklassifitseerija puhul kasutati 10-kordset ristvalideerimist. Meediatekstide klassifitseerimise põhjal selgus, et leitud mudelid suudavad paremini ennustada õigeksmõistva otsusega tekste, tauniva otsuse saanud tekstide puhul on leitud mudelite ennustusvõime madalam. Parimaks taunivat otsust ennustavaks mudeliks osutus fastText klassifitseerijaga loodud mudel, mis võtab sisendiks lemmatiseeritud meediateksti.

Töö käigus tekkinud üheks kitsaskohaks oli tasuline ligipääs meediatekstidele. DIGARI Eesti artiklite portaalist on võimalik saada paberväljaannetes ilmunud tekste, kuid suur osa meediatekste ilmub veebiväljaandena ning nõuab tasulistele artiklitele ligipääsu ostmist. Töö autor hankis magistritöö kirjutamise ajal ligipääsu Ekspress Meedia, Postimehe, Äripäeva, Õhtulehe ja Meie Maa väljaannetele. Teiseks kitsaskohaks oli artiklite veebist ülesleidmine, osa huvipakkuvaid artikleid oli veebist eemaldatud (või oli pealkirja oluliselt muudetud). Kolmandaks võiks välja tuua, et väljaanded ilmselt muudavad artikleid pärast kaebuse esitamist või pärast tauniva otsuse saamist ning esialgset artiklit ei ole võimalik enam kätte saada.

Käesolevas töös analüüsiti Pressinõukogule esitatud kaebustele vastavaid eestikeelseid meediatekste töö autorile teadaolevalt esmakordselt masinõppemeetodite abil. Tekstikaeve ja masinõppe meetodid võiksid tulevikus aidata ajakirjanikke meediatekstide kirjutamisel

hinnata, kas publitseeritav ajakirjanduslik tekst on vastavuses ajakirjanduseetika koodeksi ja eetikanormidega. Käesolevas töös tehti esimene samm selles suunas, mis näitas, et eestikeelsete meediatekstide klassifitseerimine sõnapõhiselt ei oleks ajakirjandusliku teksti jaoks piisava ennustustäpsusega ning ennustav mudel vajaks täiendamist.

Meediatekstide edasisel uurimisel võiks kasutada teemade modelleerimist. Käesoleva magistr töö edasiarendusena oleks huvitav täpsemalt välja selgitada, millistel teemadel Pressinõukogule esitatud kaebustes kirjutatakse, kas tauniva hinnangu saanud tekstides on käsitletud teemasid, mida õigeksmõistva hinnangu saanud tekstides ei ole ja vastupidi. Teine edasiarendus võiks olla erinevate ajakirjandusväljaannete põhjal väljaselgitamine, kuidas väljaandja suhtub tauniva otsuse saamisesse, kas toimetustes toimuvad arutelud, kuidas ajakirjanik reageerib tauniva otsuse saamisele, kas ja kuidas toimetuse tööd parandatakse või muudetakse.

Kasutatud kirjandus

- Aggarwal, C. C., & Zhai, C. (2012). *Mining Text Data*. © Springer Science+Business Media.
- Avaliku Sõna Nõukogu. (2021). Kasutamise kuupäev: 10.04.2022, allikas Avaliku Sõna Nõukogu: <https://www.asn.org.ee/>
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 273-297. Allikas: <https://link.springer.com/content/pdf/10.1007/BF00994018.pdf>
- Eetikaveeb. (2007). (Tartu Ülikool) Kasutamise kuupäev: 07.05.2022, allikas <https://www.eetika.ee/et/145776>
- Faktikontroll. (2022). Kasutamise kuupäev: 10.04.2022, allikas Ekspress Meedia: <https://epl.delfi.ee/artikkel/91897223/meie-pohimotted>
- fastText. (2022). Kasutamise kuupäev: 30.04.2022, allikas <https://fasttext.cc/docs/en/crawl-vectors.html>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 137-144. Allikas: <https://www.sciencedirect.com/science/article/pii/S0268401214001066>
- GDI. (2020). Allikas: Globaalne Väärinformatsiooniriski Indeks: https://disinformationindex.org/wp-content/uploads/2020/10/GDI-RiskRatingsReport_Estonian.pdf
- Hinno, R. (2018). *E-kirjade klassifitseerimine masinõppe abil. Magistritöö*. Arvutiteaduse instituut. Tartu Ülikool.
- Joachims, T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. *European Conference on Machine Learning*, (lk 137-142). Allikas: <https://link.springer.com/chapter/10.1007/BFb0026683>
- Joulin, A., Grave, E., & Bojanowski, P. M. (2016). Bag of Tricks for Efficient Text Classification. Allikas: <https://arxiv.org/pdf/1607.01759.pdf>
- Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed. draft). Allikas: <https://web.stanford.edu/~jurafsky/slp3/>
- Karu, L. (2013). *Ajakirjanduseetika koodeksi tõlgendamise vähese töökogemusega ajakirjanike poolt. Bakalaureusetöö*. Ajakirjanduse ja kommunikatsiooni instituut. Tartu Ülikool.
- Kikas, B., & Himma-Kadakas, M. (juuni 2021. a.). Ajakirjanikud ja faktikontroll. Tulevikutrikid minevikuoskuste kontekstis. *Vikerkaar*, lk 94-103.
- Kübar, E. (2006). *Meedia eneseregulatsiooni süsteemi efektiivsuse analüüs. Bakalaureusetöö*. Ajakirjanduse ja kommunikatsiooni osakond. Tartu Ülikool.
- Orasmaa, S., Petmanson, T., Tkachenko, A., Laur, S., & Kaalep, H.-J. (2016). ESTNLTK - NLP Toolkit for Estonian. Allikas: <http://www.lrec-conf.org/proceedings/lrec2016/summaries/332.html>
- Pressinõukogu. (2022). Kasutamise kuupäev: 10.04.2022, allikas Eesti Meediaettevõtete Liit: <https://meedialiit.ee/>
- Propastop. (2022). Kasutamise kuupäev: 10.04.2022, allikas <https://www.propastop.org/Reporters+Sans+Frontieres>. (3. mai 2022. a.). Allikas: <https://rsf.org/en/index?year=2022>
- scikit-learn. (2022). Kasutamise kuupäev: 26.04.2022, allikas <https://scikit-learn.org/stable/>
- Zolotov, V., & Kung, D. (2017). Analysis and Optimization of fastText Linear Text Classifier. Allikas: <https://arxiv.org/abs/1702.05531>
- VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media, Inc. Allikas: <https://jakevdp.github.io/PythonDataScienceHandbook/>
- Wardle, C., & Derakhsan, H. (2017). *INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making*. Kasutamise kuupäev: 03.03.2022, allikas <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html#>

Lisad

I. Eesti ajakirjanduseetika koodeks

1. Üldised põhimõtted.

- 1.1. Demokraatliku ühiskonna toimimise eeltingimus on kommunikatsioonivabadus. Vaba ajakirjandus on selle tingimuse saavutamise vahend ja eeldus.
- 1.2. Ajakirjandus teenib avalikkuse õigust saada tõest, ausat ja igakülgset teavet ühiskonnas toimuva kohta. Ajakirjanduse üks peamine kohustus on ühiskonnas kriitiliselt jälgida poliitilise ja majandusliku võimu teostamist.
- 1.3. Vaba ajakirjandust, kui ta täidab kehtivaid õigusakte, ei tohi mingil moel piirata ega takistada info kogumisel ja avaldamisel.
- 1.4. Ajakirjanik vastutab oma sõnade ja loomingu eest. Ajakirjandusorganisatsioon kannab hoolt selle eest, et ei ilmuks ebatäpne, moonutatud või eksitav informatsioon.
- 1.5. Ajakirjandus ei tohi oma tegevusega kellelegi tekitada põhjendamatu kannatusi, veendumata, et avalikkusel on tõesti vaja seda informatsiooni teada.
- 1.6. Poliitilist ja majanduslikku võimu ning avalikkusele olulist informatsiooni valdavaid inimesi käsitleb ajakirjandus avaliku elu tegelastena, kelle tegevuse üle on ajakirjanduse tavalisest suurem tähelepanu ja kriitika õigustatud. Samuti käsitleb ajakirjandus avaliku elu tegelastena neid, kes teenivad elatist enda isiku või loomingu eksponeerimisega.

2. Sõltumatus.

- 2.1. Ajakirjanik ei võta vastu ametikohti, soodustusi, tasu ega kingitusi, mis tekitavad seoses tema ajakirjanikutööga huvide konflikti ja võivad vähendada tema usaldusväärust.
- 2.2. Ajakirjanik, kes edastab rahandus- ja majandusalast informatsiooni, ei tohi seda levitada eraviisiliselt ega kasutada isiklikes huvides.
- 2.3. Ajakirjanik ei tohi olla kajastatava asutuse või institutsiooni teenistuses.
- 2.4. Toimetuse töötajat ei saa tööalaselt kohustada kirjutama või tegema midagi taolist, mis on vastuolus tema isiklike tõekspidamistega.

3. Ajakirjanik ja informatsiooniallikas.

- 3.1. Ajakirjanik, kogudes materjali avaldamise/edastamise jaoks, peab teatama vestluspartnerile, et ta on ajakirjanik ja millise väljaande/jaama juurest. Soovitav on teatada ka, mille jaoks informatsiooni kogutakse.
- 3.2. Ajakirjanik ei või kuritarvitada meediaga suhtlemisel kogunenud inimesi. Enne vestlust selgitatakse räägitu võimalikke tagajärgi.
- 3.3. Ajakirjanik peab rangelt kinni informatsiooniallikale antud lubadustest ja väldib lubadusi, mida ta ei suuda täita.
- 3.4. Ajakirjandusel on moraalne kohustus kaitsta konfidentsiaalseid informatsiooniallikaid.
- 3.5. Toimetus kontrollib, eelkõige kriitilise materjali korral, informatsiooni tõesust ja allikate usaldusväärust. Ka juhul kui avaldatava/edastatava materjali autoriks ei ole toimetuse töötaja, kontrollib toimetus oluliste faktide tõesust.
- 3.6. Lapsi tuleb üldjuhul intervjuuerida või temast ülevõtteid teha lapsevanema või lapse eest vastutava isiku juuresolekul või nõusolekul. Sellest reeglist võib teha erandeid, kui intervjuu kaitseb lapse huve või kui ta on niigi avalikkuse tähelepanu all.
- 3.7. Ajakirjanik hangib heli- ja pildiülevõtteid ning informatsiooni avalikult. Erandid on lubatud ülekaaluka avaliku huvi olemasolul, kui materjali pole võimalik teisiti hankida.

4. Avaldamisreeglid.

- 4.1. Uudised, arvamused ja oletused olgu selgelt eristatavad. Uudismaterjal põhinegu tõestataval ja tõenditega tagatud faktilisel informatsioonil.

- 4.2. Konflikti sisaldava materjali puhul peab ajakirjanik ära kuulama kõik osapooled.
- 4.3. Ei sobi rõhutada rahvust, rassi, religioosset või poliitilist kuuluvust ega sugu, kui sellel pole uudisväärtust.
- 4.4. Ajakirjandus ei või inimest käsitleda kurjategijana enne sellekohast kohtuotsust.
- 4.5. Enesetappude ja enesetapukatsete uudisväärtust tuleb tõsiselt kaaluda.
- 4.6. Andmeid ja arvamusi konkreetsete inimeste tervisliku (nii vaimse kui füüsilise) seisundi kohta ei avaldata, välja arvatud juhul, kui inimene on andmete avaldamisega nõus või kui niisuguste andmete avaldamist nõuab avalikkuse huvi.
- 4.7. Vanemate vaidlusi laste hooldusküsimuste üle üldjuhul ei kajastata.
- 4.8. Avaldades materjale õigusrikkumistest, kohtuasjadest ja õnnetustest peab ajakirjanik kaaluma, kas asjaosaliste identifitseerimine on tingimata vajalik ja milliseid kannatusi võib see asjaosalistele põhjustada. Ohvreid ja alaealisi kurjategijaid üldjuhul avalikkuse jaoks ei identifitseerita.
- 4.9. Inimese eraelu puutumatus rikkuvaid materjale avaldatakse vaid juhul, kui avalikkuse huvid kaaluvad üles inimese õiguse privaatsusele.
- 4.10. Tsitaatide, fotode, heli- ja videomaterjali kasutamisel originaalist erinevas kontekstis tuleb olla ettevaatlik. Eksitus-ohtlik montaaž, samuti heli või pildi moonutus olgu varustatud asjakohase märke või teadustusega.
- 4.11. Fotod, fotode allkirjad, pealkirjad, juhtlauseid ega saatetutvustused ei või auditooriumi eksitada.
- 4.12. Välisautori poolt loodud ajakirjandusliku materjali sisu, konteksti ja avaldamisaega ei ole hea muuta autori teadmise ja nõusolekuta.

5. Vastulause.

- 5.1. Kui kellegi kohta avaldatakse tõsiseid süüdistusi, tuleks talle võimaluse korral pakkuda kommentaari võimalust samas numbris või saates.
- 5.2. Vastulausega on õigus parandada avaldatud materjalis esinevaid faktivigu ja tsitaate. Vastulauseks ei saa nõuda rohkem ruumi/aega, kui oli kasutada kriitikaks. Vastulause tuleb avaldada viivitusteta ja märgataval kujul, ilma toimetusepoolsete hinnanguliste kommentaarideta.
- 5.3. Ebaõige informatsiooni ilmumise korral tuleb avaldada parandus.

6. Reklaam.

- 6.1. Reklaam ja suhtekorraldusmaterjal olgu auditooriumi jaoks selgelt eristatud ajakirjanduslikust tekstist/pildist.
- 6.2. Ajakirjanikud ja alalised välisautorid ei edasta samas kanalis reklaamteateid ega kirjuta oma väljaandes oma nime all kommertstekste.
- 6.3. Toodet või kaubamärki esitletakse ajakirjanduslikus materjalis vaid juhul, kui see on põhjendatud.
- 6.4. Tarbijale suunatud ajakirjandusliku materjali puhul peab auditooriumile selgitama, kuidas vastavate toodete valik tehti ning kuidas tooteid testiti.

Heaks kiidetud EALLi üldkoosolekul 29.01.1998

II. Tabelid

Tabel 12. Pressinõukogule esitatud kaebused aastate 2002-2021 lõikes.

Kaebuse esitamise aasta	Kaebuste arv aastas
2002	14
2003	33
2004	32
2005	45
2006	35
2007	24
2008	45
2009	54
2010	42
2011	67
2012	49
2013	56
2014	51
2015	74
2016	84
2017	87
2018	86
2019	82
2020	54
2021	67
Kokku	1081

Tabel 13. Esitatud kaebused kõigi väljaannete lõikes.

Väljaande nimi	Kaebusi	Osakaal, %
Postimees	169	15,6
Eesti Ekspress	102	9,4
Eesti Päevaleht	102	9,4
Delfi	97	9
Õhtuleht	82	7,6
Äripäev	55	5,1
Kanal2	54	5
ERR	48	4,4
TV3	35	3,2
Pärnu Postimees	23	2,1
muu väljaanne	20	1,9
Lääne Elu	20	1,9
Saarte Hääl	19	1,8
Meie Maa	16	1,5
Võrumaa Teataja	15	1,4
Hiiu Leht	13	1,2

Maaleht	13	1,2
Sakala	11	1
Den za Dnjom	11	1
Harju Elu	11	1
MK-Estonia	9	0,8
Põhjarannik	9	0,8
Virumaa Teataja	9	0,8
Järva Teataja	8	0,7
Delovõje Vedomosti	6	0,6
Narvskaja Gazeta	6	0,6
Elu24	6	0,6
ERR portaal	5	0,5
Gorod	5	0,5
LõunaLeht	5	0,5
Lääne-Harju Ekspress	5	0,5
Kroonika	5	0,5
Pealinn	5	0,5
Estonija	4	0,4
Vooremaa	4	0,4
Müürileht	4	0,4
Koit	4	0,4
Sirp	3	0,3
Tallinna TV	3	0,3
Sillamäeskii Vestnik	3	0,3
Oma Saar	3	0,3
Kesknädal	3	0,3
Molodjoz Estonii	3	0,3
Vesti Dnja	2	0,2
Viru Prospekt	2	0,2
Paldiski Linnaleht	2	0,2
Kuku Raadio	2	0,2
Valgamaalane	2	0,2
Pirita	2	0,2
Just!	2	0,2
Raplamaa Sõnumid	2	0,2
Stolitsa	2	0,2
Õpetajate Leht	2	0,2
Vesti Nedeli	2	0,2
Sõnumitooja	2	0,2
Meditšiiniuudised	2	0,2
Elva Postipoiss	1	0,1
Pere&Kodu	1	0,1
BNS	1	0,1
Harjumaa	1	0,1
Lõuna-Eesti Postimees	1	0,1

Krengolmski Prospekt	1	0,1
Sky.ee	1	0,1
Kose Teataja	1	0,1
Nädaline	1	0,1
Puutepunktid	1	0,1
Nelli Teataja	1	0,1
Naisteleht	1	0,1
Kultuur ja Elu	1	0,1
Eesti Kirik	1	0,1
Saldo	1	0,1
Videvik	1	0,1
Tartu Linnaleht	1	0,1
Panoraam	1	0,1
Virumaa Nädalaleht	1	0,1
Vali Uudised	1	0,1
Vesti	1	0,1
Linnaleht	1	0,1
Kokku	1081	100

Tabel 14. Ajakirjanduseetika koodeksipunktide mainimine kaebuste lõikes.

Koodeksipunkte ühes kaebuses	Kaebuste arv
0	699
1	261
2	104
3	7
4	7
5	2
7	1
Kokku	1081

Tabel 15. Ajakirjanduseetika koodeksi mainimine koodeksipunktide lõikes.

Koodeksipunkt	Mainimiste arv
1.2	1
1.4	97
1.5	20
1.6	1
2.3	1
3.1	5
3.2	4
3.3	5
3.5	25
3.6	2
3.7	2
4.1	68

4.2	121
4.3	4
4.4	11
4.6	2
4.7	2
4.8	6
4.9	7
4.10	5
4.11	69
4.12	6
5.1	53
5.2	12
5.3	5
6.1	1
Kokku	535

Tabel 16. Ajakirjanduseetika koodeksipunktid õigeksmõistvates ja taunivates otsustes (andmestikus Kaebused_1).

Koodeksipunkte ühes kaebuses	Kaebuste arv
0	529
1	259
2	104
3	7
4	7
5	2
7	1
Kokku	909

Tabel 17. Ajakirjanduseetika koodeksipunktide mainimine õigeksmõistvates ja taunivates otsustes (andmestikus Kaebused_1).

Koodeksipunkt	Mainimiste arv	Osakaal, %
1.2	1	0,2
1.4	97	18,2
1.5	20	3,8
1.6	1	0,2
2.3	1	0,2
3.1	3	0,6
3.2	4	0,8
3.3	5	0,9
3.5	25	4,7
3.6	2	0,4
3.7	2	0,4
4.1	68	12,8
4.2	121	22,7

4.3	4	0,8
4.4	11	2,1
4.6	2	0,4
4.7	2	0,4
4.8	6	1,1
4.9	7	1,3
4.10	5	0,9
4.11	69	12,9
4.12	6	1,1
5.1	53	9,9
5.2	12	2,3
5.3	5	0,9
6.1	1	0,2
Kokku	533	100

Tabel 18. Ajakirjanduseetika koodeksipunktide mainimine taunivates otsustes (andmestik *Kaebused_1*).

Koodeksipunkt	Mainimiste arv	Osakaal, %
1.2	1	0,2
1.4	96	18,4
1.5	19	3,6
1.6	0	0
2.3	1	0,2
3.1	3	0,6
3.2	4	0,8
3.3	5	1
3.5	23	4,4
3.6	2	0,4
3.7	2	0,4
4.1	67	12,9
4.2	118	22,6
4.3	4	0,8
4.4	11	2,1
4.6	2	0,4
4.7	2	0,4
4.8	4	0,8
4.9	7	1,3
4.10	5	1
4.11	69	13,2
4.12	6	1,2
5.1	52	10
5.2	12	2,3
5.3	5	1
6.1	1	0,2
Kokku	521	100

Tabel 19. Õigeksmõistvate ja taunivate otsuste jaotus kõigi väljaannete lõikes (andmestikus Kaebused_1).

Väljaande nimi	tauniv	õigeksmõistev	kokku	tauniv, %	õigeksmõistev, %
Postimees	63	80	143	44,1	55,9
Eesti Ekspress	39	56	95	41,1	58,9
Eesti Päevaleht	47	44	91	51,6	48,4
Delfi	49	39	88	55,7	44,3
Õhtuleht	24	45	69	34,8	65,2
Äripäev	21	29	50	42	58
Kanal2	23	22	45	51,1	48,9
TV3	14	14	28	50	50
ERR	11	17	28	39,3	60,7
Pärnu Postimees	13	7	20	65	35
Saarte Hääl	5	13	18	27,8	72,2
Lääne Elu	8	9	17	47,1	52,9
Meie Maa	9	6	15	60	40
muu väljaanne	7	7	14	50	50
Hiiu Leht	8	5	13	61,5	38,5
Võrumaa Teataja	9	4	13	69,2	30,8
Maaleht	4	7	11	36,4	63,6
Harju Elu	5	5	10	50	50
Sakala	3	7	10	30	70
Põhjarannik	1	8	9	11,1	88,9
Den za Dnjom	4	5	9	44,4	55,6
MK-Estonia	5	4	9	55,6	44,4
Virumaa Teataja	2	6	8	25	75
Järva Teataja	2	5	7	28,6	71,4
Elu24	3	3	6	50	50
Kroonika	2	3	5	40	60
Lääne-Harju Ekspress	3	2	5	60	40
Pealinn	1	3	4	25	75
Gorod	3	1	4	75	25
Delovõje Vedomosti	2	2	4	50	50
Müürileht	1	3	4	25	75
Narvskaja Gazeta	1	2	3	33,3	66,7
Sirp	0	3	3	0	100
Tallinna TV	2	1	3	66,7	33,3
Vooremaa	1	2	3	33,3	66,7
Estonija	2	1	3	66,7	33,3
Stolitsa	2	0	2	100	0
Vesti Dnja	2	0	2	100	0
Raplamaa Sõnumid	2	0	2	100	0

Vesti Nedeli	1	1	2	50	50
Õpetajate Leht	0	2	2	0	100
Oma Saar	2	0	2	100	0
Molodjoz Estonii	0	2	2	0	100
LõunaLeht	0	2	2	0	100
Koit	2	0	2	100	0
Kuku Raadio	0	2	2	0	100
Puutepunktid	1	0	1	100	0
Vesti	1	0	1	100	0
ERR portaal	0	1	1	0	100
Eesti Kirik	0	1	1	0	100
Elva Postipoiss	1	0	1	100	0
Harjumaa	0	1	1	0	100
Kesknädal	1	0	1	100	0
Virumaa Nädalaleht	1	0	1	100	0
Viru Prospekt	0	1	1	0	100
Krengolmski Prospekt	1	0	1	100	0
Valgamaalane	1	0	1	100	0
Vali Uudised	1	0	1	100	0
Pirita	1	0	1	100	0
Linnaleht	1	0	1	100	0
Lõuna-Eesti Postimees	0	1	1	0	100
Sõnumitooja	0	1	1	0	100
Meditšiiniuudised	1	0	1	100	0
Sillamäeskii Vestnik	0	1	1	0	100
Saldo	1	0	1	100	0
Nädaline	0	1	1	0	100
Pere&Kodu	0	1	1	0	100
BNS	1	0	1	100	0
Kokku	421	488	909		

IV. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Anne-Liis Rämson**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

Pressinõukogule esitatud kaebuste otsuste ennustamine masinõppe abil,

mille juhendaja on Kairit Sirts,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Anne-Liis Rämson
14.05.2022