

# Koneoppimiseen perustuva sentimenttianalyysi tuotteiden kehittämisen tukena

TURUN YLIOPISTO  
Tietotekniikan laitos  
TkK-tutkielma  
Tietotekniikka  
Huhtikuu 2026  
Oskari Arvo

TURUN YLIOPISTO  
Tietotekniikan laitos

OSKARI ARVO: Koneoppimiseen perustuva sentimenttianalyysi tuotteiden kehittämisen tukena

TkK-tutkielma, 41 s.  
Tietotekniikka  
Huhtikuu 2026

---

Tässä tutkielmassa tarkastellaan, miten koneoppimiseen perustuvaa sentimenttianalyysia hyödynnetään tuotteiden kehittämisen tukena. Lähtökohtana on, että verkossa kertyvä käyttäjien tuottama sisältö, kuten tuotearvostelut ja sosiaalisen median julkaisut, sisältää runsaasti arvokasta tietoa asiakkaiden kokemuksista, odotuksista ja kehitysehdotuksista. Tutkielma on toteutettu kirjallisuuskatsauksena, jossa tarkastellaan erityisesti käytettyjä koneoppimismenetelmiä, aineistolähteitä ja niiden vaikutusta menetelmävalintoihin sekä eri lähestymistapojen vahvuuksia ja rajoitteita tuotekehityksen näkökulmasta. Katsauksen perusteella tuotekehityksessä hyödynnetään laajasti perinteisen koneoppimisen luokittelijoita sekä syväoppimis- ja transformer-pohjaisia malleja ja analyysi kohdistuu yhä useammin aspektitasolle, jotta havainnot voidaan liittää suoraan yksittäisiin tuoteominaisuuksiin ja niiden kehittämiseen. Tutkielmassa havaittiin, että menetelmien onnistuminen riippuu vahvasti datan laadusta, esikäsittelystä, piirre-esityksestä ja arviointiasetelmasta. Lisäksi käytännön hyödyn saavuttaminen ei useimmiten ole sidoksissa yksittäiseen parhaaseen malliin. Keskeisiä haasteita ovat tulosten vertailtavuus eri tutkimusten välillä, yleistettävyyys eri alustoihin ja tuoteryhmiin sekä mallien tulkittavuus, joka on ratkaisevaa, jotta analyysin tuloksia voidaan luotettavasti hyödyntää tuotekehityksen päätöksenteossa.

Asiasanat: sentimenttianalyysi, perinteinen koneoppiminen, syväoppiminen, tuotekehitys, verkkokauppa, sosiaalinen media

# Sisällys

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Taustaa</b>	<b>4</b>
2.1	Sentimenttianalyysi . . . . .	4
2.2	Koneoppivat menetelmät . . . . .	9
2.2.1	Perinteinen koneoppiminen . . . . .	16
2.2.2	Syväoppiminen . . . . .	20
<b>3</b>	<b>Kirjallisuuskatsaus sentimenttianalyysin hyödyntämisestä tuotekehityksessä</b>	<b>28</b>
3.1	Koneoppimiseen perustuvan sentimenttianalyysin soveltaminen tuotekehityksessä . . . . .	28
3.2	Käytetyt aineistot ja niiden vaikutukset menetelmiin . . . . .	31
3.3	Keskeiset tulokset ja vertailu . . . . .	32
<b>4</b>	<b>Pohdinta</b>	<b>35</b>
<b>5</b>	<b>Yhteenveto</b>	<b>39</b>
	<b>Lähdeluettelo</b>	<b>42</b>

# Kuvat

1.1	Kirjallisuuskatsauksen lähteiden hakuprosessi visualisoituna. . . . .	3
2.1	Kappaleessa käyty sentimenttianalyysi tai mielipidelouhinta visualisoituna. . . . .	10
2.2	Koneoppimiseen perustuvan sentimenttianalyysin prosessi, joka etenee datan keruusta esikäsittelyyn ja piirteiden poiminnan kautta mallinnukseen, arviointiin ja suorituskyvyn mittaamiseen. . . . .	17
2.3	Kaksiluokkaisen SVM-luokittelun havainnollistus kolmella mahdollisella hypertasolla. . . . .	19
2.4	Esimerkki CNN-mallinnuksesta tekstinluokittelussa. . . . .	23
2.5	Esimerkki ketjumaisesta LSTM-mallista. . . . .	25
2.6	Esimerkki BERT-mallin kouluttamisesta. . . . .	27
4.1	Yhdysvaltalaisten jälleenmyyjien verkkokaupan markkinaosuus 2022 visualisoituna. . . . .	37

# 1 Johdanto

Sentimenttianalyysi on keskeinen luonnollisen kielen käsittelyn (NLP) menetelmä, jonka avulla tietokoneet voivat tunnistaa tekstissä ilmaistuja mielipiteitä, tunteita ja asenteita [1], [2], [3], [4]. Sen nopea kasvu liittyy suoraan sosiaalisen median, blogien, keskustelupalstojen ja verkkokauppojen yleistymiseen, joka on lisännyt huomattavasti ihmisten tuottaman tekstimuotoisen datan määrää digitaalisessa muodossa [4], [5], [6]. Tuotearvioissa ja muissa teksteissä esitetyt mielipiteet tarjoavat arvokasta tietoa tuotteiden ja palveluiden valmistajille, jälleenmyyjille ja palveluntarjoajille, minkä vuoksi sentimenttianalyysi on noussut merkittäväksi tutkimusalueeksi [1], [5]. Sitä käytetään laajasti sovelluksissa, kuten asiakaspalautteiden analysoinnissa, sosiaalisen median seurannassa ja tuotearvosteluissa [1], [6]. Arvostelut voivat toimia innovatiivisten ideoiden lähteenä ja tarjota arvokasta tietoa uusien tuotteiden suunnitteluun ja parantamiseen [7].

Sentimenttianalyysin varhaiset menetelmät, kuten sanastopohjaiset menetelmät, eivät kykene tunnistamaan kielen vivahteita tai merkityseroja tehokkaasti. Koneoppimismenetelmät sen sijaan oppivat monimutkaisia kielellisiä piirteitä suoraan datasta, mikä parantaa sentimenttianalyysin suorituskykyä [6]. Tässä tutkielmassa keskitytään eri koneoppimismenetelmien käyttöön sentimenttianalyysissa. Tutkielmassa ei oteta huomioon sanastopohjaisia menetelmiä, mutta otetaan huomioon hybridimenetelmät, jos ne käsittelevät koneoppivia menetelmiä.

Tutkielman päätavoitteena on selvittää, mitä koneoppimismenetelmiä käytetään sentimenttianalyysissa tuotekehityksen tukena ja mistä näihin menetelmiin tarvittava data saadaan. Lisäksi tavoitteena on tarkastella, miten datan ominaisuudet vaikuttavat menetelmän valintaan sekä mitä vahvuuksia ja heikkouksia eri menetelmiin liittyy. Näiden pohdintojen perusteella muodostettiin seuraavat tutkimuskysymykset:

**Tutkimuskysymys 1:** Mitä eri koneoppimismenetelmiä käytetään sentimenttianalyysissa tuotekehityksen tukena ja mihin sovelluskohteisiin niitä hyödynnetään?

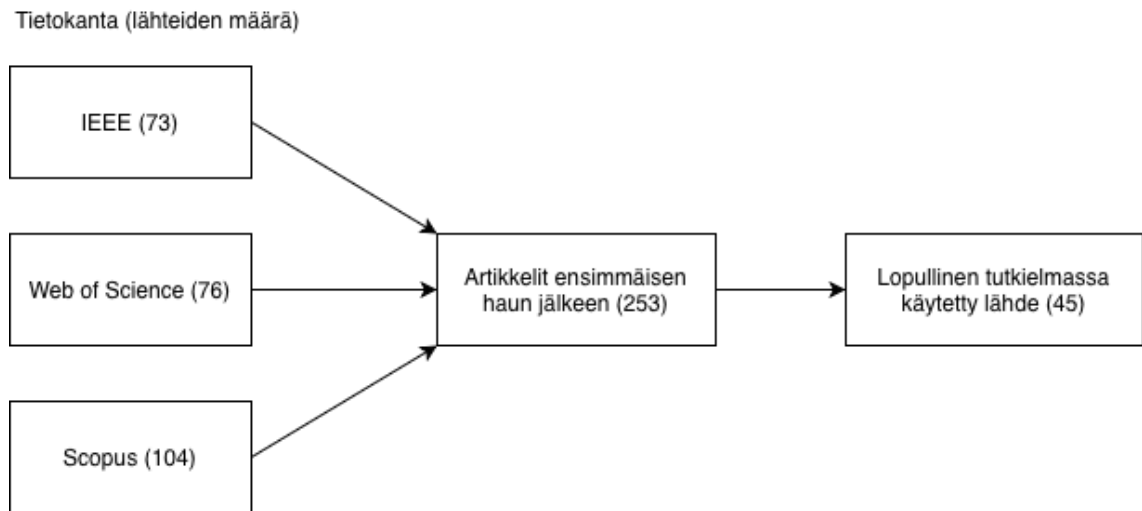
**Tutkimuskysymys 2:** Mistä saadaan dataa ja miten data vaikuttaa siihen mitä koneoppimismenetelmää käytetään?

**Tutkimuskysymys 3:** Mitkä ovat eri sentimenttianalyysin koneoppimismenetelmien vahvuudet ja heikkoudet tuotekehityksen tukena?

Tämä kandidaatintutkielma on toteutettu kirjallisuuskatsauksena. Tässä tutkielmassa käytetyt lähteet ovat pääosin peräisin kolmesta tietokannasta: IEEE, Web of Science ja Scopus. Lisäksi lähteitä täydennettiin käymällä läpi valittujen julkaisujen lähdeluetteloiden lähteitä, jotta hakuun saatiin mukaan myös aiheen kannalta olennaisia lisäjulkaisuja. Varsinainen kirjallisuuskatsaus rajattiin vuosina 2021–2025 julkaistuihin tutkimuksiin ajantasaisen tiedon varmistamiseksi.

Kirjallisuuskatsauksen hakuprosessissa käytettiin ensisijaisesti seuraavaa hakulauseketta: ("product development" OR "customer needs") AND ("natural language processing" OR NLP OR "text mining" OR BERT OR GPT) AND "sentiment analysis" ja sen variaatioita rajaamaan hakua. Haun tuloksena IEEE:stä löytyi 73 artikkelia, Web of Sciencesta 76 artikkelia ja Scopusesta 104 artikkelia, jolloin hakutulosten kokonaismääräksi muodostui 253 artikkelia. Näistä valittiin sopivimmat artikkelit katsaukseen.

Työn teoreettisessa taustassa hyödynnetään myös vanhempaa kirjallisuutta keskeisten käsitteiden ja mallien määrittelyä ja taustatiedoa varten käytettiin hakulauseita kuten "sentiment analysis" ja "sentiment analysis" AND "machine learning". Kuvassa 1.1 on visualisoitu kirjallisuuskatsauksen lähteiden hakuprosessi:



Kuva 1.1: Kirjallisuuskatsauksen lähteiden hakuprosessi visualisoituna.

Tutkielman alussa esitellään työn teoreettinen tausta: mitä sentimenttianalyysi on ja mitä sen eri osa-alueet sisältävät, mitä koneoppiminen tarkoittaa sekä miten perinteinen koneoppiminen ja syväoppiminen eroavat toisistaan. Tämän jälkeen esitetään kirjallisuuskatsauksen tulokset. Tulosten tarkastelu keskittyy siihen, mitä koneoppimismenetelmiä käytetään sentimenttianalyysissa tuotekehityksen tukena, mihin sovelluskohteisiin näitä käytetään, mistä tarvittava data on peräisin, miten datan ominaisuudet vaikuttavat menetelmävalintaan sekä mitkä ovat eri menetelmien vahvuudet ja heikkoudet. Sitten käsitellään tutkitun alan pohdinta ja haasteita. Lopuksi esitetään tutkielman yhteenveto ja alan tulevaisuuden näkymiä.

## 2 Taustaa

### 2.1 Sentimenttianalyysi

Sentimenttianalyysi tutkimusalana yhdistää tiedonhaun, luonnollisen kielen käsittelyn (NLP) ja tekoälyn [2], [8]. Tämän vuoksi samankaltaisesta asiasta käytetään erilaisia nimityksiä, kuten sentimenttianalyysi, mielipidelouhinta tai subjektiivisuusanalyysi. Vaikka termeillä on pieniä eroavaisuuksia, ne viittaavat pääosin samaan tutkimusalueeseen [2]. Termiä sentimenttianalyysi käytetään yleisesti NLP:n yhteydessä ja sillä tarkoitetaan laskennallista menetelmää, jossa tunnistetaan tekstissä ilmaistuja mielipiteitä, tunteita ja asenteita [2], [3], [4]. Yksi sen keskeisistä tehtävistä on mielipiteiden tunnistaminen ihmisten tuottamasta tekstistä ja sen luokittelu esimerkiksi positiiviseksi, negatiiviseksi tai neutraaliksi [9], [10].

Yleisesti ottaen sentimenttianalyysia voidaan tarkastella kolmella eri tasolla: dokumentti-, lause- ja aspektipohjaisella tasoilla. Sentimenttianalyysin varhaisessa vaiheessa analysoitiin kokonaisia dokumentteja ja luokiteltiin niiden yleinen sentimentti positiiviseksi tai negatiiviseksi [9], [11], [12]. Dokumenttitaso on näistä yksinkertaisin ja se soveltuu hyvin silloin, kun teksti käsittelee vain yhtä aiheita tai kohdetta, kuten elokuva- tai tuotearvosteluissa [9], [11]. Usein teksti sisältää kuitenkin useita mielipiteitä tai tunteita eri aiheista, eivätkä tunteet rajoitu vain yhteen kohteeseen. Tämän vuoksi sentimenttianalyysin tarkkuutta alettiin parantaa siirtymällä

lausetasolle, koska yksittäisten lauseiden analysointi mahdollistaa subjektiivisten ja objektiivisten ilmausten tarkemman tunnistamisen [11].

Lausetasolla tavoitteena on selvittää, ilmaiseeko lause positiivisen, negatiivisen vai neutraalin mielipiteen [8] ja lisäksi arvioida sen subjektiivisuus [9]. Esimerkiksi lause "Tänään on iloinen päivä" ilmaisee myönteisen tunteen, joten se on subjektiivinen väite, joka voidaan luokitella tunnearvon perusteella. Sen sijaan tosiasioita kuvaavat lauseet ovat objektiivisiä, kuten esimerkiksi lause "Tänään on aurinkoinen päivä", joka ei suoraan ilmaise tunnetta ja siksi se tyypillisesti luokitellaan objektiiviseksi väitteeksi [6]. Samoin kuin dokumenttitasolla, niin lausetasolla oletetaan, että vain yksi mielipide ilmenee koko lauseessa. Esimerkiksi lauseessa "Uuden älypuhelimien näytön koko on hyvä, mutta sen akun kesto on huono" tunnearvo kohdistuu kahteen eri kohteeseen. Lausetasolla näitä ei huomioida, vaan arvioidaan koko lauseen sentimentti [11]. Näin siis lausetason lähestymistavat eivät aina onnistu havaitsemaan mielipiteitä tai tunteita, jotka kohdistuvat tiettyyn entiteettiin tai sen ominaisuuksiin.

Tämän rajoitteen vuoksi kehitettiin aspektipohjainen lähestymistapa, jossa sentimentti kohdistetaan tiettyyn entiteettiin ja sen ominaisuuksiin [11]. Verrattuna dokumentti- ja lausetasoon, aspektipohjainen sentimenttianalyysi keskittyy tekstin yksityiskohtiin ja pyrkii tunnistamaan mielipiteet nimenomaan kohteen eri piirteistä [2], [9], [13]. Sen tehtävä koostuu kolmesta vaiheesta: kohteen tunnistamisesta, sen ominaisuuksien erottelusta sekä ominaisuuksien luokittelemisesta positiiviseksi, negatiiviseksi tai neutraaleiksi [9]. Tämän avulla voidaan tuottaa kokonaiskuva kohteesta tai entiteetistä sekä tilastoja sen ominaisuuksiin liittyvistä mielipiteistä, mikä tekee aspektipohjaisesta tasosta sentimenttianalyysin hienojakoisimman muodon [8], [11]. Esimerkiksi tuotearvosteluissa itse tuote on yleensä kohde tai entiteetti, kun taas kaikki siihen liittyvät asiat, kuten esimerkiksi hinta, laatu tai muut ominaisuudet, ovat tuotteen aspekteja [2]. Tätä havainnollistaa lause "Uuden älypuhelimien

näytön koko on hyvä, mutta sen akun kesto on huono", jossa arvioidaan entiteetin (älypuhelimien) kahta aspektia: näytön kokoa ja akun kestoa. Näytön kokoon kohdistuva tunnearvo on positiivinen, kun taas akun keston kohdistuva tunnearvo on negatiivinen [11]. Dokumentti- tai lausetason sentimenttianalyysissä näitä kahta arviota ei eroteltaisi, vaan lauseelle annettaisiin yksi tunnearvo kokonaisuutena. Aspektitaso on erityisen hyödyllinen tuotekehityksessä, koska se kertoo mihin ominaisuuteen palaute kohdistuu [14], [15].

Tällä hetkellä suurin osa sentimenttianalyysin menetelmistä voidaan jakaa kolmeen pääkategoriaan: sanastopohjaisiin menetelmiin, koneoppimiseen perustuviin menetelmiin sekä hybridi-lähestymistapoihin [9], [12], [13], [16]. Sanasto- tai tietopohjaiset menetelmät olivat ensimmäisiä sentimenttianalyysissä käytettyjä lähestymistapoja ja ne voidaan jakaa kahteen alaryhmään: sanakirja- ja korpuspohjaisiin menetelmiin [13], [16]. Sanakirjapohjaisessa menetelmässä tekstin sentimentti määritetään käyttämällä valmiita sanastoja, kuten SentiWordNet [17], jossa kullekin sanan merkitykselle on määritelty positiivinen tai negatiivinen arvo. WordNet [18] toimii usein näiden sanastojen pohjana. Tekstin kokonaissentimentti muodostetaan näiden arvojen perusteella [9], [13], [16]. Korpuspohjaisessa menetelmässä ei käytetä valmista sanastoa, vaan se rakennetaan siemensanaston avulla tilastollisesti analysoimalla laajempaa tekstiaineistoa [13], [16], [19]. Sanasto- tai tietopohjaisissa menetelmissä teksti luokitellaan tunneluokkiin tunnistamalla siitä tunnetta ilmaisevia sanoja, kuten esimerkiksi "iloinen", "surullinen" tai "vihainen" [20]. Sanastopohjaisten menetelmien etuina ovat niiden yksinkertaisuus, kustannustehokkuus sekä se, että ne eivät vaadi opetusdataa toisin kuin koneoppimiseen perustuvat menetelmät [20]. Menetelmien haasteena on kuitenkin sanastoresurssien kattavuus ja laatu, sillä valmiit sanastot eivät huomioi esimerkiksi slangia tai sarkasmia ja sanojen merkitys voi muuttua asiayhteydestä riippuen. Tämän vuoksi sama sana voi olla yhdessä aiheessa positiivinen, mutta toisessa neutraali tai jopa negatiivinen,

mikä heikentää menetelmän tarkkuutta erityisesti kontekstiriippuvaisissa teksteissä, kuten sosiaalisen median sisällössä [12], [16]. Esimerkiksi sanastopohjainen menetelmä pystyy usein tunnistamaan yksinkertaisia tunnesanoja, kuten sanan "iloinen" ja luokittelemaan lauseen "Tänään oli iloinen päivä" positiiviseksi. Sen sijaan se voi tulkita virheellisesti lauseen "Tänään ei ollut iloinen päivä ollenkaan", jos menetelmä ei huomioi kieltosanaa, koska sanastopohjaiset menetelmät eivät aina osaa ottaa kieltä oikein huomioon ilman erikseen määriteltyä sääntöä [20]. Sanastopohjaisten menetelmien suorituskyky riippuu siis vahvasti sanastojen laadusta ja niiden soveltuvuudesta analysoitavaan aineistoon [13], [20].

Koneoppimiseen perustuvat tai tilastolliset menetelmät, ovat suosittuja tekstien sentimentin luokittelussa [20]. Nämä menetelmät mallintavat sanojen esiintymistä ja niiden välisiä tilastollisia suhteita, joiden perusteella sentimenttiä voidaan luokitella. Koneoppimiseen perustuvat menetelmät voidaan jakaa kahteen ryhmään: perinteisiin malleihin ja syväoppimismalleihin. Perinteisillä malleilla tarkoitetaan klassisia koneoppimismenetelmiä, kuten Naive Bayes (NB) -luokittelijaa tai Support Vector Machine (SVM) -menetelmää [9], [13]. Näissä menetelmissä hyödynnetään koneoppimisalgoritmeja sentimentin luokitteluun ja piirteiden määrittely sekä poiminta ovat keskeisessä roolissa [9]. Piirteet voivat olla esimerkiksi sanoja ja tietoa niiden käytöstä tekstissä. Koneoppimismenetelmien tarkkuus riippuu hyvin vahvasti valittujen piirteiden laadusta ja soveltuvuudesta, eli kuinka hyvin valitut piirteet kykenevät luokittelemaan tekstiä [8], [12]. Tyypillisiä piirteitä ovat bag-of-words-pohjaiset menetelmät, kuten term frequency–inverse document frequency (TF-IDF) ja n-grammit, sekä sanaupotukset (engl. word embeddings), kuten Word2Vec ja GloVe [12]. Perinteisissä koneoppimismenetelmissä piirteet määritellään ja poimitaan joko käsin tai hyödyntämällä piirteentalinta- ja muodostusmenetelmiä. Piirteiden suunnittelu ja poiminta voi kuitenkin olla työlästä, sillä hyvän piirteentalinta- ja poimintasuunnittelun suunnittelu vaatii huomattavaa teknistä osaamista ja toimialakohtaista asian-

tuntemusta ja piirteenmuodostus on usein yksi prosessin aikaa vievimmistä vaiheista [6], [8], [13], [21]. Perinteiset koneoppimismenetelmät perustuvat ennalta määriteltyihin piirteisiin, minkä vuoksi niiden kyky tavoittaa tekstin merkityksiä on rajallinen verrattuna syväoppimismalleihin. Huolellisesti suunnitellut piirteet voivat parantaa tuloksia, mutta menetelmät eivät mallinna laajempaa kontekstia yhtä hyvin [10], [20]. Yksittäiset sanat eivät yleensä riitä kertomaan luotettavasti, onko teksti positiivinen vai negatiivinen. Näiden menetelmien suorituskyky paranee tyypillisesti analysoitavan tekstimäärän kasvaessa. Tämän takia ne voivat toimia hyvin esimerkiksi dokumenttitasolla, mutta usein huonommin pienemmissä yksiköissä, kuten yksittäisissä lauseissa [20].

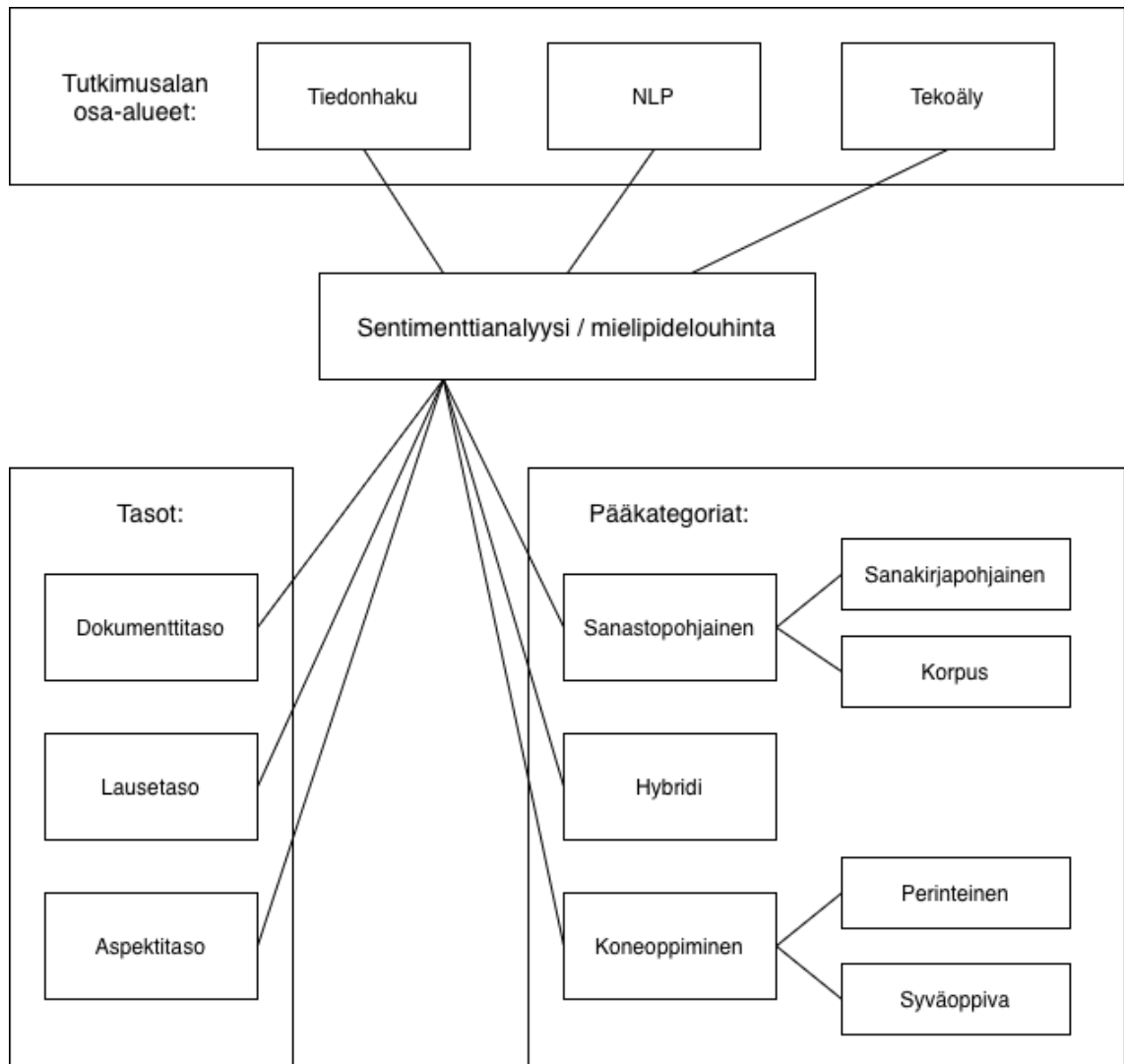
Perinteiset koneoppimispohjaiset menetelmät eivät aina anna riittävän tarkkoja tuloksia, minkä vuoksi syväoppimismallit ovat nousseet keskeisiksi luonnollisen kielen käsittelyn sovelluksissa. Niiden tarjoamien etujen vuoksi syväoppimista pidetään tärkeänä tutkimusalueena [6], [9]. Dataa syntyy nykyään valtavia määriä, eivätkä kaikki perinteisen koneoppimisen menetelmät aina pysty käsittelemään sitä tehokkaasti ja luotettavasti. Syväoppimismallit voivat hyödyntää suuria aineistoja tehokkaasti, koska ne oppivat monimutkaisia piirteitä ilman ennalta määriteltyjä piirteitä toisin kuin perinteiset menetelmät. Tämä voi parantaa suorituskykyä tai tarkkuutta [6], [10]. Syväoppiminen perustuu monikerroksisiin neuroverkkoihin, joissa hyödynnetään piilotettuja kerroksia monimutkaisten rakenteiden oppimiseen [13]. Syväoppimismallien etu on, että ne oppivat itse tunnistamaan tärkeät piirteet datasta ilman, että ihminen joutuu määrittelemään niitä käsin. Tämä tekee niistä hyödyllisiä erityisesti siksi, että ne kykenevät hyödyntämään suuria datamääriä ja laskentatehoa esimerkiksi sentimenttianalyysissa [6], [13], [21]. Haasteena on kuitenkin se, että ne vaativat suuria määriä dataa oppiakseen toimimaan luotettavasti [6]. Yksi syväoppimismallien haasteista on hallusinaatiot, joissa mallit voivat olla varmoja myös silloin, kun niiden ennusteet ovat virheellisiä. Lisäksi syväoppimismalleilla on

niin sanottu musta laatikko -luonne, mikä tarkoittaa heikkoa selitettävyyttä, minkä vuoksi on usein vaikeaa tai mahdotonta tarkasti selittää, miksi malli on tehnyt tietyn luokittelun, kuten esimerkiksi arvioinut lauseen positiiviseksi tai negatiiviseksi. Koska syväoppimismallit sisältävät valtavan määrän parametreja, niiden opettaminen on usein aikaa vievää. Hyvän suorituskyvyn saavuttaminen edellyttää yleensä myös tehokkaita laskentaresursseja, kuten GPU-laitteistoa ja suurta RAM-muistin määrää [6]. Sentimenttianalyysiin voidaan käyttää useita erilaisia syväoppimismalleja, kuten Convolutional Neural Network (CNN), Deep Neural Network (DNN) ja Recurrent Neural Network (RNN) tai transformerit [12], [13]. Syväoppiminen jäljittelee ihmisaivojen oppimisprosessia, joka pohjautuu neuroverkkojen käsitteeseen. Syväoppimismallien resurssivaatimuksista huolimatta ne ovat usein tarkempia ja tehokkaampia kuin perinteiset menetelmät. Viime vuosien aikana syväoppimisesta on tullut keskeinen lähestymistapa koneoppimisessa ja se on osoittanut parempaa suorituskykyä kuin perinteiset koneoppimismenetelmät useilla sovellusalueilla mukaan lukien luonnollisen kielen käsittelyssä, jonka alle sentimenttianalyysi kuuluu [12].

Hybridisessä lähestymistavassa yhdistetään sanastopohjaisia menetelmiä koneoppimiseen perustuviin lähestymistapoihin [9]. Sanastot ovat keskeisessä roolissa suurimmassa osassa näistä menetelmistä [13]. Tämä tutkimus keskittyy koneoppimiseen perustuviin menetelmiin, eli perinteisiin ja syväoppiviin sentimenttianalyysin lähestymistapoihin. Kuvassa 2.1 on visualisoitu kappaleessa käyty sentimenttianalyysi tai mielipidelouhintaa, sekä sen alalajit ja niiden väliset suhteet.

## 2.2 Koneoppivat menetelmät

Koneoppimismallit ovat herättäneet paljon huomiota viime vuosina [10]. Kuten äskeisessä kappaleessa mainittiin, koneoppimiseen perustuvat menetelmät voidaan jakaa kahteen ryhmään: perinteisiin malleihin ja syväoppimismalleihin [13].



Kuva 2.1: Sentimenttianalyysin tutkimusalan osa-alueet (tiedonhaku, NLP ja tekoäly), tasot (dokumentti-, lause- ja aspektitaso) sekä keskeiset menetelmäkategoriat (sanastopohjaiset, koneoppimiseen perustuvat ja hybridi-lähestymistavat) ja niiden väliset suhteet.

Sentimenttianalyysissa käytetään yleisesti kahta koneoppimistekniikkaa: ohjattua ja ohjaamatonta oppimista [9], [16], [22]. Ohjatut menetelmät perustuvat valmiiksi merkittyyntä opetusdataan ja edellyttävät yleensä suurta määrää merkittävää dataa, jotta mallit voivat oppia tehokkaasti. Ohjaamattomassa oppimisessä vastaavaa merkittävää aineistoa ei ole saatavilla, minkä vuoksi sitä käytetään erityisesti tilanteissa, joissa luokiteltua dataa on vaikea hankkia [16], [22].

Koneoppimiseen perustuvissa menetelmissä merkittyyden aineistojen rajallinen saatavuus on keskeinen haaste. Ohjatun oppimisen mallit edellyttävät yleensä suuria määriä merkittävää dataa mallin parametrien optimoimiseksi, minkä vuoksi suorituskyky on usein vahvasti sidoksissa opetusdatan määrään ja laatuun. Tämä korostuu erityisesti uusilla aihealueilla, joilla saatavilla oleva merkittävä data on usein vähäistä [12]. Koneoppimiseen perustuvat mallit ovat lisäksi usein aihealueeseen sidonnaisia. Ne toimivat parhaiten silloin, kun testidata muistuttaa koulutusdataa, mutta niiden suorituskyky heikkenee merkittävästi, kun niitä sovelletaan koulutusdatasta poikkeavaan aihealueeseen. Tällöin malli joudutaan usein kouluttamaan uudelleen tai hienosäätämään, jotta tulokset säilyvät luotettavina [12].

Valtaosa maailmanlaajuisesti saatavilla olevasta datasta on jäsentymätöntä, kuten tekstiä, puhetta, ääntä ja videoita. Tekstipohjainen data voi sisältää myös lyhenteitä, kirjoitusvirheitä ja huonosti jäsenneityjä lauseita, mikä voi heikentää mallin suorituskykyä [3], [8], [12]. Tämän vuoksi tarvitaan esikäsittelyvaihe, joka toteutetaan yleensä luonnollisen kielen käsittelyn menetelmillä [12]. Jäsentymättömän tekstidatan käsittely edellyttää useita vaiheita, sillä teksti voi olla peräisin monista eri lähteistä ja esiintyä erilaisissa muodoissa. Tämänkaltaisen datan tehokas käsittely on osaltaan edistänyt luonnollisen kielen käsittelyn alojen, kuten tiedonhaun ja tiedonlouhinnan kehittymistä. Datan hankinta ja esikäsittely ovat keskeisiä tekstilouhinnan ja sentimenttianalyysin alatehtäviä [3]. Sentimenttianalyysi edellyttää tekstimuotoisen opetusdatan esikäsittelyä, jossa aineistosta poistetaan kohinaa ja sitä muokataan mallille sopivaan muotoon ennen mallin kouluttamista [13]. Vaikka tutkimus usein keskittyy mallien ja menetelmien kehittämiseen, esikäsittelyvaiheella on merkittävä vaikutus lopulliseen suorituskykyyn ja asianmukaisella esikäsittelyllä mallin tarkkuutta voidaan parantaa [8]. Tekstidatan esikäsittelyyn sisältyy tyypillisesti normalisointi, tokenisointi ja kohinan vähentäminen [12]. Tokenisointi jakaa tekstin yksittäisiksi tokeneiksi, jotka voidaan tämän jälkeen esittää esimerkiksi bag-

of-words-mallin tai n-grammien avulla. NLP-sovelluksissa tokenisointi luokitellaan usein osaksi esikäsitteilyä, vaikka se vaikuttaakin suoraan mallinnuksessa käytettävään piirre-esitykseen. Esikäsitteilyn järjestys ja valitut menetelmät vaikuttavat lopputulokseen ja eri koneoppimismallit voivat hyötyä erilaisista esikäsitteilytavoista [12]. Näiden toimenpiteiden tavoitteena on varmistaa, että piirteiden poimintavaiheessa hyödynnetään vain tehtävän kannalta olennaista tietoa [12].

Useimmat perinteiset koneoppimiseen perustuvat mallit noudattavat kaksivaiheista menetelmää. Ensimmäisessä vaiheessa dokumenteista tai muista tekstiyksiköistä poimitaan ja valitaan käsin määritellyjä piirteitä [10], [16]. Toisessa vaiheessa nämä piirteet syötetään luokittimelle, joka tekee ennusteen [10]. Ohjattujen koneoppimismenetelmien keskeinen piirre on niiden riippuvuus valmiiksi merkitystä opetusdatasta. Perinteiset koneoppimismenetelmät edellyttävät lisäksi, että asiantuntijat valitsevat käytettävät piirteet, jotta malli pystyisi tekemään mahdollisimman tarkkoja ja yleistettäviä luokituksia. Mallin suorituskyky riippuu vahvasti näiden piirteiden laadusta ja siitä, kuinka hyvin ne kuvaavat luokiteltavaa tekstiä [8], [12].

Perinteisen koneoppimisen piirteet määritellään ja poimitaan joko käsin tai erilaisia piirteervalinta- ja muodostusmenetelmiä hyödyntäen [12], [13]. Tyypillisiä tekstin piirteitä ovat bag-of-words-menetelmät, kuten term frequency–inverse document frequency (TF-IDF) ja n-grammit [10], [12], [13], [16]. Koska ohjattujen menetelmien tehokkuus riippuu vahvasti käytettyjen piirteiden laadusta, piirteemuodostuksessa hyödynnetään usein bag-of-words-mallin laajennuksia, kuten TF-IDF tai n-grammeja. Näiden menetelmien tavoitteena on tuottaa kuvaavampia ja paremmin yleistäviä piirteitä kuin yksinkertaiset sanapohjaiset esitykset [2].

Bag-of-words (BOW) -malli on yksi suosituimmista esitystavoista asioiden luokittelussa [23]. BOW-mallissa sanajärjestyksellä ei ole merkitystä, joten esimerkiksi "iloinen päivä" ja "päivä iloinen" tuottavat identtisen piirre-esityksen, koska ne sisältävät samat sanat yhtä monta kertaa. BOW-mallin tai vektoriavaruusesityksen

esittivät alun perin informaationhakututkijat tekstikorpusten kuvaamiseen. Se on yksinkertainen tapa muuntaa jäsentymätön teksti rakenteelliseksi dataksi käsittelemällä sitä sana sanalta ja sivuuttamalla kielioppi. Sanojen painojen laskemiseen on useita menetelmiä, kuten term frequency–inverse document frequency (TF-IDF) [24]. TF-IDF on tilastollinen mittari, joka kuvaa kuinka tärkeä jokin sana on tietylle dokumentille dokumenttikokoelmassa tai -korpuksessa [13], [22], [24]. Mittari huomioi sanan esiintymistiheyden kohdedokumentissa sekä sen esiintymistiheyden muissa korpuksen dokumenteissa [13], [25]. Termien esiintymistiheyttä verrataan käänteiseen dokumenttitiheyteen, joka mittaa, kuinka monessa dokumentissa sana esiintyy koko korpuksessa [22], [25]. Mitä useammin sana esiintyy kohdedokumentissa ja mitä harvemmin se esiintyy muissa dokumenteissa, sitä tärkeämmäksi sana arvioidaan [13]. Näin TF-IDF-menetelmä muuntaa mielivaltaisen pituiset dokumentit kiinteän mittaisiksi lukulistoiksi [25]. N-grammit on menetelmä, jossa tarkastellaan n-peräkkäistä sanaa annetusta tekstistä. Tämä ottaa huomioon sanojen järjestyksen ja niiden välisen yhteyden paremmin kuin yksittäisten sanojen tarkastelu. Sentimentianalyysissä n-grammit auttavat analysoimaan tekstin tai dokumentin sentimenttiä. Unigram viittaa n-grammiin, jonka koko on 1, bigram kokoon 2, trigram kokoon 3 ja niin edelleen. Esimerkiksi "Tänään", "oli", "iloinen", "päivä" ovat unigrammeja ja "Tänään oli", "oli iloinen", "iloinen päivä" ovat taas bigrammeja [22]. Kaksivaiheiseen lähestymistapaan liittyy useita merkittäviä haasteita. Koska se perustuu käsin määriteltyihin piirteisiin, vaaditaan runsaasti asiantuntijatyötä niiden suunnitteluun ja arviointiin, jotta mallilta saadaan hyvä suorituskyky. Tämä vahva riippuvuus aihealuekohtaisesta asiantuntemuksesta rajoittaa samalla menetelmän sovellettavuutta uusiin tehtäviin ja heikentää sen yleistettävyyttä. Lisäksi ennalta määritellyt piirteet rajoittavat mallin kykyä hyödyntää suuria opetusmääriä tehokkaasti, koska malli ei pysty oppimaan uusia piirteitä suoraan datasta [10].

Neuroverkkoihin perustuvia menetelmiä on tutkittu käsin määriteltyihin piirteisiin liittyvien rajoitusten vähentämiseksi [10]. Syväoppimismallien etuna on, etteivät ne ole riippuvaisia manuaalisesta piirteenpoiminnasta, sillä ne pystyvät oppimaan sopivat piirteet suoraan datasta, mikä vähentää aihealueen asiantuntemuksen tarvetta [12]. NLP-tehtävissä, kuten sentimenttianalyysissä, tämän mahdollistaa erityisesti sanaopetusmenetelmien käyttö, sillä ne muuntavat tekstin jatkuviksi vektoreiksi, jolloin käsin määriteltyjä piirteitä ei tarvita [10]. Sanaopotukset ovat piirteenoppimisen tekniikka, jossa jokainen sana muunnetaan tiiviiksi jatkuviksi vektoreiksi siten, että samankaltaiset sanat saavat keskenään samankaltaiset esitykset. Piirteiden oppiminen voidaan toteuttaa neuroverkoilla tai muilla matemaattisilla menetelmillä [6], [13]. Yleisesti käytettyjä sanaopetusmenetelmiä ovat Word2Vec [26] ja GloVe [27], joita on saatavilla myös valmiiksi koulutettuina malleina [12]. Word2Vec on laajalti käytetty neuroverkkopohjainen menetelmä sanavektorien oppimiseen. Se hyödyntää matalaa neuroverkkoa oppiakseen esittämään samankaltaisissa yhteyksissä esiintyvät sanat toisiaan muistuttavina vektoreina. Menetelmän tuottamat vektoriesitykset ovat rakenteeltaan sellaisia, että niille voidaan suorittaa myös yksinkertaisia laskutoimituksia [6], [10]. GloVe on laajalti käytetty sanaopetusmenetelmä, joka perustuu sanojen laskennalliseen yhteisesiintymistietoon. Se oppii sanavektorit hajottamalla sanojen yhteisesiintymismatriisin osiin ja sen etuna on erityisesti nopea opetettavuus suurilla aineistoilla, sillä sen toteutus on mahdollista rinnakkaistaa [6], [10].

Koneoppimismallin koulutuksen jälkeen sen suorituskykyä arvioidaan erillisellä aineistolla, jota kutsutaan testijoukoksi. Testijoukon avulla mitataan mallin yleistämiskykyä, eli sen kykyä suoriutua uusista syötteistä, joita se ei ole kohdannut koulutusvaiheessa [21]. Rajallisella koulutusdatalla malli voi oppia myös datan kohinaa, joka esiintyy koulutusaineistossa, mutta ei toistu uudessa testijoukossa, vaikka testijoukko olisi peräisin samasta jakaumasta [28]. Tällöin malli oppii koulutusdatansa

liian tarkasti ja menettää kyvyn yleistyä, mikä johtaa ylisovittamiseen [29]. Ylisovittamista pyritään vähentämään esimerkiksi varhaisella koulutuksen pysäyttämällä sekä regularisointimenetelmillä, kuten L1- ja L2-regularisoinnilla [28]. Alioppiminen taas tarkoittaa, että malli ei opi koulutusdataa riittävän hyvin, minkä seurauksena se suoriutuu heikosti sekä koulutusaineistolla, että testijoukolla [29].

Ohjattujen koneoppimismallien suorituskyvyn arvioinnissa hyödynnetään sekaannusmatriisiin perustuvia mittareita [22]. Sekaannusmatriisin avulla arvioidaan koneoppimismallin toimivuutta vertaamalla havaittuja ja ennustettuja luokkia ja sen keskeisiä käsitteitä ovat True Positive (TP), False Positive (FP), True Negative (TN) ja False Negative (FN). True Positive kuvaa niiden tapausten määrää, joissa teksti sisältää positiivisen sentimentin ja malli on luokitellut sen oikein positiiviseksi. False Positive puolestaan tarkoittaa tapauksia, jotka on luokiteltu virheellisesti positiiviseksi, vaikka ne eivät todellisuudessa ole positiivisia. Vastaavasti True Negative viittaa tapauksiin, jotka ovat negatiivisia ja jotka on luokiteltu oikein negatiiviseksi, kun taas False Negative tarkoittaa tapauksia, jotka on luokiteltu virheellisesti negatiiviseksi, vaikka ne eivät todellisuudessa ole negatiivisia. [10], [22].

Perustuen sekaannusmatriisista saatuihin arvoihin voidaan laskea mallin suorituskykyä arvioivia mittareita [10], [22]. Accuracy (suom. tarkkuus) ja error rate (suom. virhetiheys) ovat yleisimmin käytettyjä mittareita mallin laadun arvioimisessa [10]. Accuracy lasketaan oikein luokitellut tapaukset suhteessa kaikkien tapausten kokonaismäärään [22]. Error rate puolestaan lasketaan väärin luokitellut tapaukset suhteessa kaikkien tapausten kokonaismäärään [22]. Nämä määritellään yhtälöissä:

$$\text{Accuracy} = \frac{TP + TN}{N}, \quad \text{Error rate} = \frac{FP + FN}{N}, \quad (2.1)$$

missä  $N$  on tapausten kokonaismäärä.

Näiden lisäksi on muun muassa precision (suom. tarkkuuspositiivisuus), recall (suom. herkkyys) ja F1-score (suom. F1-mitta) [10], [22]. Precision kertoo, kuin-

ka luotettavia positiiviset ennusteet ovat. Se on oikein tunnistettujen positiivisten tapausten osuus kaikista tapauksista, jotka malli luokittelee positiivisiksi [22]. Tämä määritellään yhtälössä:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.2)$$

Recall kertoo, kuinka hyvin malli tunnistaa positiiviset tapaukset. Se on oikein tunnistettujen positiivisten tapausten osuus kaikista todellisista positiivisista tapauksista [22]. Tämä määritellään yhtälöissä:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.3)$$

F1-score yhdistää precisionin ja recallin yhdeksi luvuksi. Sitä käytetään arvioimaan mallin kokonaissuorituskykyä silloin, kun molemmat mittarit halutaan huomioida tasapuolisesti [22]. Tämä määritellään yhtälössä:

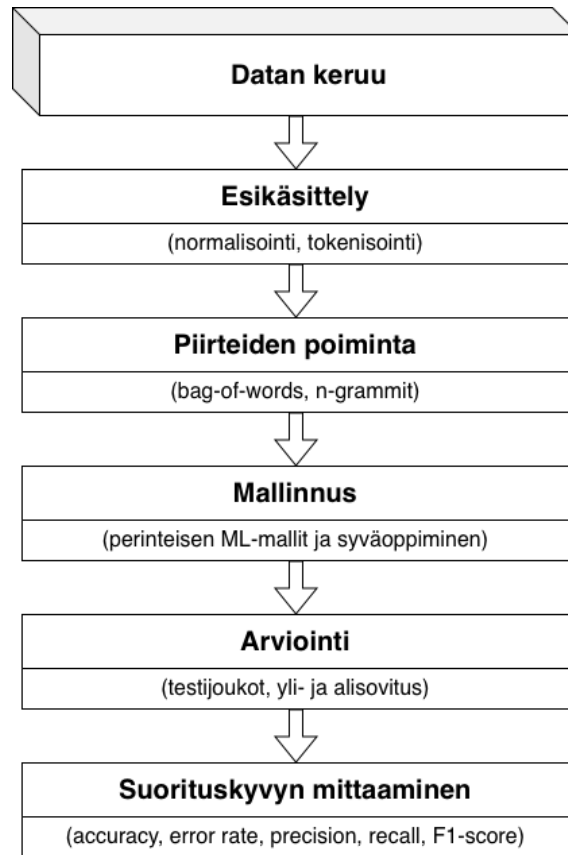
$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.4)$$

Nämä mittarit ovat keskeisiä mallin laadun arvioimisessa [10].

Kuvassa 2.2 esitetään luvussa käsitellyt koneoppimiseen perustuvan sentimenttiansalyysin vaiheet visualisoituna.

### 2.2.1 Perinteinen koneoppiminen

Yksi varhaisimmista ja merkittävimmistä sentimenttiansalyysin ja koneoppimisen yhdistävistä tutkimuksista on Pangin ja Leen työ 2000-luvun alusta [30]. He hyödynsivät IMDb-elokuva-arvosteluja ja sovelsivat muun muassa Naive Bayes -luokittelijaa, SVM:ää ja maksimientropiamenetelmää dokumenttitason sentimenttiansalyysissa. Tätä tutkimusta pidetään yhtenä keskeisenä perustana myöhemmälle sentimenttiansalyysin tutkimukselle [3]. Myöhemmin McDonald et al. esittivät rakenteellisen



Kuva 2.2: Koneoppimiseen perustuvan sentimenttianalyysin prosessi, joka etenee datan keruusta esikäsittelyyn ja piirteiden poiminnan kautta mallinnukseen, arviointiin ja suorituskyvyn mittaamiseen.

mallin, joka luokittelee sentimentin samanaikaisesti sekä lause- että dokumenttitasolla koneoppimisen avulla [31]. Malli perustui oletukseen siitä, että lauseiden ja dokumenttien sentimenttiluokat ovat keskenään riippuvaisia. Kokeet suoritettiin tuotearvosteluilla, jotka oli kerätty Amazon.com-sivustolta. Tulokset osoittivat, että ehdotettua mallia voidaan laajentaa luokittelemaan sentimenttiä useammalla tasolla kuin pelkästään lause- ja dokumenttitasolla [3].

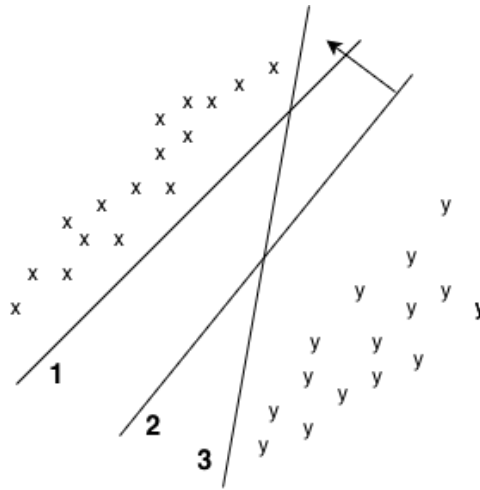
Perinteinen koneoppiminen voidaan jakaa ohjattuihin ja ohjaamattomiin oppimismenetelmiin [6]. Kirjallisuuden perusteella sentimenttianalyysissa ohjattuun oppimiseen perustuvia menetelmiä on selvästi enemmän kuin ohjaamattoman oppimisen menetelmiä, sillä ohjattujen mallien tutkimus ja käyttö on ollut pitkään hallit-

sevassa asemassa [6]. Suosituttuja ohjatun koneoppimisen menetelmiä sentimenttianaalyyssissa ovat muun muassa Naive Bayes ja Support Vector Machine -menetelmät.

Yksi yksinkertaisimmista ja yleisimmin käytetyistä perinteisistä koneoppimismalleista on Naive Bayes (NB) -luokitin, joka on Bayesin teoreemaan perustuva todennäköisyyspohjainen koneoppimismenetelmä. Sitä käytetään usein tekstinluokittelutehtävissä, kuten sentimenttianaalyyssissä [1], [12], [16], [22]. Sentimenttianaalyyssissä NB-menetelmä arvioi todennäköisyyden, että teksti kuuluu johonkin sentimenttiluokkaan (esimerkiksi positiiviseen, negatiiviseen tai neutraaliin), analysoimalla yksittäisten sanojen esiintymistodennäköisyyksiä [1]. Bayesin teoreemaa ja ehdollisia todennäköisyyksiä hyödyntämällä voidaan laskea posterioritodennäköisyys, eli todennäköisyys sille, että dokumentti kuuluu johonkin luokkaan tietyn piirrejoukon ehdolla. NB-menetelmässä tämä posteriori arvioidaan käyttämällä dokumentissa havaittujen sanojen frekvenssejä ja niitä vastaavia tilastoja, jotka on opittu koulutusaineistosta [12], [16]. Menetelmä perustuu niin sanottuun "naiiviin" oletukseen, jonka mukaan tietyn piirteen esiintyminen oletetaan olevan riippumaton muista piirteistä luokan ehdolla. Vaikka tämä oletus on yksinkertaistava, se toimii käytännössä yllättävän hyvin monissa tilanteissa [1], [12], [16]. Vaikka NB-menetelmä on rakenteeltaan yksinkertainen, se on erittäin nopea, helposti toteutettava ja toimii luotettavasti myös pienissä aineistoissa erityisesti silloin, kun riippumattomuusoletus pitää paikkansa [1]. Malli sopii hyvin myös suurempiin datakokonaisuuksiin sen tehokkaan laskennan ja suoraviivaisen käytännön toteutuksen ansiosta [12]. Menetelmän etuna on sen kyky käsitellä kohinaista dataa varsin tehokkaasti, mutta jos piirteet eivät ole riippumattomia, sen suorituskyky voi heikentyä [1].

Support Vector Machine (SVM) -menetelmän perusideana on analysoida dataa ja löytää piirreavaruudesta erotuspinta, eli hypertaso, joka erottaa eri luokkiin kuuluvia datapisteitä, maksimoimalla marginaalin niiden välillä [12], [16], [22]. SVM pysyy muodostamaan epälineaarisia päätösrajoja muuntamalla aineiston korkeamman

ulottuvuuden piirreavaruuteen, jossa luokat voidaan erottaa lineaarisesti hypertason avulla [16]. Tämä lähestymistapa tunnetaan yleisesti nimellä "kernel trick". Tekstidata soveltuu erityisen hyvin SVM-luokitteluun tekstin harvan rakenteen vuoksi [16]. Yksi yleinen esimerkki tämänkaltaisesta datasta on bag-of-words-malli. Koska SVM-menetelmät käsittelevät vain numeerisia vektoreita, tekstipohjaiset arviot on ensin muutettava numeeriseen muotoon. Tämän jälkeen vektoreiden arvoja voidaan vielä käsitellä niin, että piirteet pysyvät keskenään vertailukelpoisina [22]. Menetelmää hyödynnetään laajasti erilaisissa luokittelutehtävissä, kuten sentimenttiansalysissä ja esimerkiksi arvostelujen luokittelussa niiden sisällön perusteella [12], [16]. SVM-menetelmien suorituskyky voi kuitenkin heikentyä, jos data sisältää paljon kohinaa. Lisäksi erityisesti suurten aineistojen kohdalla mallin opettaminen voi olla aikaa vievää [12]. Kuvassa 2.3 havainnollistetaan kaksiluokkainen SVM-luokittelu kolmella mahdollisella hypertasolla.



Kuva 2.3: Esimerkiksi kaksiluokkaisessa luokittelutehtävässä SVM etsii juuri sen hypertason, joka tuottaa suurimman marginaalin ja siten optimaalisen erottelun luokkien välille [12], [22]. Seuraavassa kuvassa on esitetty luokat x ja y sekä kolme mahdollista hypertasoa: 1, 2 ja 3. Kuvan perusteella hypertaso 2 on paras vaihtoehto, koska se on kauimpana lähimmistä datapisteistä ja näin muodostaa suurimman marginaalin luokkien välille [16]. Tätä havainnollistetaan kuvassa. Lähde: Mukailtu [16]

Toinen perinteisen koneoppimisen oppimistapa on ohjaamaton oppiminen. Yleisesti ottaen tämän luokan mallit toimivat ohjaamattomasti. Näiden mallien onnistuneeseen kouluttamiseen tarvitaan yleensä suuria datamääriä [2]. Latent Dirichlet Allocation (LDA) [25] on 2000-luvun alun tutkimuksessa esitelty ohjaamattoman koneoppimisen menetelmä, jota käytetään tekstissä esiintyvien selkeiden ja piilevien aiheiden tunnistamiseen, joita voidaan hyödyntää piirteiden muodostamisessa [32]. Se on kolmitasoinen hierarkkinen bayesilainen malli ja sitä on käytetty monissa tutkimuksissa dokumenttien piilevien aiheiden löytämiseen [5], [25], [32]. LDA:n taustalla oleva oletus on generatiivinen dokumenttimalli, jossa dokumentit syntyvät aiheiden ja sanojen sekoitetun jakauman perusteella [16], [25], [32]. LDA:ssa kukin dokumentti nähdään seoksena aiheista, jotka voivat olla tuottaneet kyseisen dokumentin [2], [25]. LDA:ta voidaan käyttää piirteiden löytämiseen jostain aiheesta, kuten esimerkiksi tuotteen ominaisuuksien löytämiseen tekstistä [32], [33]. LDA:n heikkous on, että sen tuottamat aiheet ovat nimeämättömiä, joten niitä ei voi suoraan yhdistää tiettyihin ominaisuuksiin tai piirteisiin. Vaikka aiheeseen liittyvät sanat voivat joskus paljastaa sen merkityksen, kaikki aiheet eivät ole helposti tulkittavia [2]. Myös perinteinen LDA-algoritmi olettaa dokumentin olevan bag-of-words-mallin mukainen, joten se ei huomioi tekstin sanojen järjestystä tai niiden välisiä suhteita [2], [8]. Tämän seurauksena jotkin aiheet voivat olla niin epäselviä, ettei niitä ole mielekästä tulkita, mikä heikentää LDA:n tulkittavuutta [2]. Jokaisen sanan aihealueen määrittäminen dokumentissa tapahtuu riippumattomasti edeltävistä ja seuraavista sanoista [5].

### 2.2.2 Syväoppiminen

Syväoppimisen merkitys korostui erityisesti vuoden 2015 jälkeen, kun LeCun et al. [21] julkaisussaan esittivät syväoppimisen vakiintuneen lähestymistavan monille koneoppimisen sovelluksille. Syväoppimisen hyödyntäminen NLP-tehtävissä on lisään-

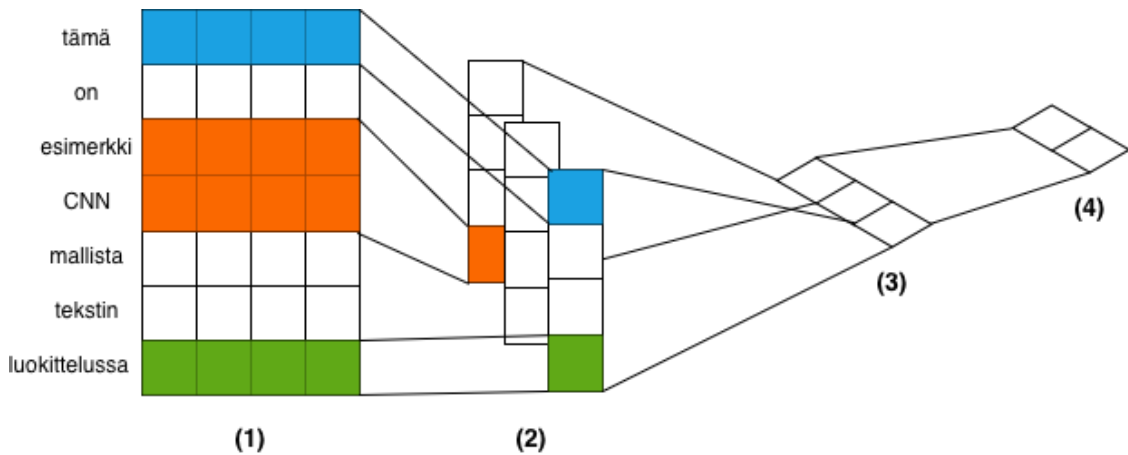
tynyt merkittävästi viime vuosien aikana. Neuroverkkopohjaisten mallien nopea kehitys on johtanut huomattaviin parannuksiin monissa luonnollisen kielen käsittelyn sovelluksissa, kuten sentimenttianalyysissa [8].

Syväoppiminen on tekoälyn osa-alue, jonka mallit rakentuvat useista keinotekoisia neuroneita sisältävistä kerroksista. Mallien taustalla on alun perin ihmisaivoista inspiroitunut ajatus siitä, että neuronien väliset yhteydet voivat vahvistua tai heikentyä kokemuksen myötä [9], [12]. Neuroverkon perusyksikkö on neuronin ja verkko muodostuu näiden neuroneiden kerroksista [13]. Syväoppiminen perustuu neuroverkkoihin, jotka koostuvat useista peräkkäisistä kerroksista. Toisin kuin vähemmän kerroksia sisältävät matalat mallit, syväoppimismallit koostuvat useista piilokerroksista ja verkon kerrosmäärä määrittää sen syvyyden [9], [12]. Syvissä neuroverkoissa (DNN) syöte- ja ulostulokerroksen välissä on yksi tai useampi piilokerros, joiden yksiköitä kutsutaan perinteisesti piiloyksiköiksi tai piilokerroksen neuroneiksi [6], [9], [13], [21]. Aiemmissä kerroksissa olevien neuronien ulostulot toimivat myöhempien kerrosten syötteinä, mikä mahdollistaa tiedon käsittelyn vaiheittaisena ketjuna [16]. Piirre-esitysten oppimisella tarkoitetaan menetelmiä, joissa mallille voidaan antaa data sellaisenaan ja sen odotetaan oppivan itse ne piirre-esitykset, joita tarvitaan esimerkiksi luokittelutehtävissä. Syväoppimismenetelmät ovat piirre-esitysten oppimismenetelmiä tehtävien ratkaisemista varten [21]. Piilokerrokset käsittelevät syötettä lineaaristen muunnosten ja aktivointifunktioiden avulla, minkä seurauksena verkko oppii vähitellen monimutkaisempia ja epälineaarisia piirre-esityksiä. Näiden esitysten oppiminen tapahtuu takaisinkytkentäalgoritmien (backpropagation) avulla, jolloin verkko kykenee havaitsemaan rakenteita myös jäsentymättömästä datasta [6], [12], [13]. Luokittelutehtävissä, kuten sentimenttianalyysissa näiden muunnosten seurauksena luokat ovat usein paremmin ja viimeisessä kerroksessa lähes lineaarisesti eroteltavissa [10], [21]. Tämän kerroksittaisen esitysten oppimisen ansiosta neuroverkko pystyy tuottamaan lopullisen luokittelutuloksen annetun syötteen perusteella

[9], [12], [21]. Suosittuja syväoppimismenetelmiä sentimenttianalyysissa ovat muun muassa CNN, RNN ja transformer-pohjaiset esikoulutetut mallit.

Konvoluutioneuroverkkomallit (CNN) kuuluvat neuroverkkojen luokkaan ja ovat saavuttaneet merkittävää menestystä ja edistysellisyttä tietokonenäön ja kuvankäsittelyn alueella [6], [8]. CNN on syväoppimisen eteenpäin syöttävä (feed-forward) neuroverkkotyyppi, jonka arkkitehtuuria on inspiroinut ihmisen näköjärjestelmän toiminta [12], [13]. Eteenpäin syöttävät neuroverkot ovat yksinkertaisimpia neuroverkkopohjaisia malleja. Tekstiä voidaan käsitellä sanaupotusten avulla, joissa sanat esitetään vektorimuodossa esimerkiksi menetelmillä, kuten Word2Vec tai GloVe. Nämä esitykset toimivat neuroverkon syötteenä. Näiden summa tai keskiarvo muodostaa tekstin esityksen, joka syötetään yhden tai useamman eteenpäin syöttävän kerroksen läpi, minkä jälkeen luokittelu tehdään neuroverkon viimeisessä kerroksessa [10]. CNN:t osoittautuivat huomattavasti helpommiksi opettaa ja paremmin yleistyviksi kuin täysin kytketyt verkot, mikä johti niiden käyttöönottoon jo aikana, jolloin neuroverkot eivät olleet erityisen suosittuja [21]. Nykyisin CNN-malleja hyödynnetään muun muassa konenäössä, suosittelujärjestelmissä ja luonnollisen kielen käsittelyssä [13]. NLP-tehtävissä CNN:tä on käytetty menestyksekkäästi esimerkiksi sentimenttianalyysiin, konekääntämiseen ja kysymyksiin vastaamiseen [1], [8]. Tyyppillinen CNN koostuu useista peräkkäisistä kerroksista. Malli alkaa syötekerroksella, joka vastaanottaa verkolle annettavan datan. NLP-tehtävissä teksti muunnetaan sana- tai merkkikohtaisiksi vektorimuotoisiksi piirre-esityksiksi, jotka järjestetään matriisiksi [8]. Muodostuva matriisi toimii neuroverkon syötteenä ja on tulkittavissa kaksiuotteisena tensorina (kuva 2.4, kohta 1). Syötettä käsitellään konvoluutio- ja poolauskerroksissa, joiden tehtävänä on poimia syötteestä olennaisia piirteitä ja tiivistää syntynyttä esitystä [6], [12], [21]. Konvoluutiokerrokset oppivat suodattimien avulla havaitsemaan paikallisia rakenteita, kuten tekstissä esiintyviä sanayhdistelmiä, jotka voivat olla hyödyllisiä esimerkiksi sentimentin tunnistamisessa (kuva 2.4,

kohta 2) [1]. Poolauskerrokset pienentävät piirre-esityksen kokoa, mikä nopeuttaa laskentaa ja lisää mallin kestävyttä kohinaa ja pieniä vääristymiä vastaan (kuva 2.4, kohta 3) [12], [13]. Konvoluutio- ja poolauskerrosten tuottama piirre-esitys välitetään lopuksi täysin kytketyille kerroksille, joita käytetään luokitteluun [6], [13]. Täysin kytketyt kerrokset hyödyntävät opittuja piirteitä luokittelussa ja tuottavat lopullisen ennusteen ulostulokerroksessa (kuva 2.4, kohta 4) [12], [13]. CNN ei käsittele tekstiä sana kerrallaan, vaan tunnistaa usean sanan muodostamia rakenteita, joista se etsii tehtävän kannalta tärkeitä piirteitä. Tämän vuoksi se on tehokas havaitsemaan sentimenttiä ilmentäviä ilmauksia. CNN-mallit ovat myös NLP-tehtävissä laskennallisesti kevyitä verrattuna muihin syväoppimismalleihin, mikä tekee niistä sopivia suurten tekstiaineistojen käsittelyyn [1]. Kuvassa 2.4 havainnollistetaan CNN-mallinnus tekstinluokittelussa.

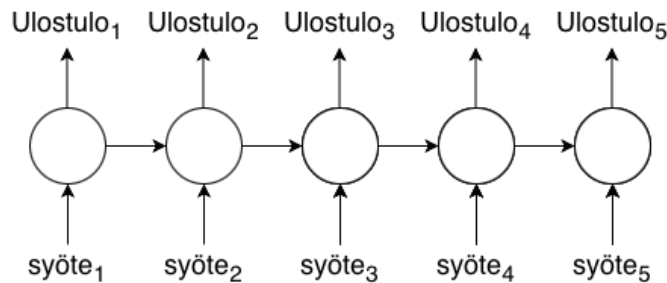


Kuva 2.4: CNN-mallinnus tekstinluokittelussa, jossa teksti muunnetaan vektorieksityksiksi ja käsitellään konvoluutio- ja poolauskerroksissa piirteiden poimimiseksi, minkä jälkeen luokittelu tehdään täysin kytketyissä kerroksissa. Numerot 1-4 viittaavat tekstissä kuvattuihin vaiheisiin. Lähde: Mukailtu [10]

CNN:t oppivat havaitsemaan rakenteita sijainnin perusteella, kun taas toistuvat neuroverkot oppivat tunnistamaan rakenteita ajan perusteella [10]. Toistuvat neuroverkot (RNN) ovat neuroverkkoja, joissa neuronien väliset yhteydet muodostavat suunnatun silmukan. Tämä silmukka mahdollistaa aiempien syötteiden tallen-

tumisen piilotiloihin, minkä ansiosta verkon nykyinen ulostulo riippuu sekä uudesta syötteestä, että aiemmin tallennetusta tilasta [12], [13]. RNN käsittelee sarjadataa askel kerrallaan siten, että verkon piilokerrokset ylläpitävät piilotiloja, jotka toimivat lyhytkestoisina muisteina aiemmista syötteistä. Tämän mekanismin ansiosta RNN:t pystyvät oppimaan sarjojen sisäisiä riippuvuuksia ja hyödyntämään kontekstia esimerkiksi seuraavan sanan ennustamisessa tai sentimentin luokittelussa [6], [12]. RNN-malleja käytetään järjestyksellisten tietojen, kuten tekstin tai puheen käsitteilyyn. Ne tarkastelevat syötettä sanajonona ja hyödyntävät sanojen välisiä riippuvuuksia esimerkiksi tekstiluokittelussa [10], [12]. RNN-verkkoja hyödynnetään monissa sarjallisen datan sovelluksissa, kuten konekääntämisessä, puheentunnistuksessa ja videoiden toimintojen tulkinnessa [6]. Yksinkertaiset RNN:t toimivat kuitenkin usein heikosti, koska ne eivät käsittele pitkiä riippuvuuksia hyvin ja kärsivät gradientin katoamisesta tai räjähtämisestä [10], [12]. Gradientti kuvaa, kuinka paljon mallin painoja tulee muuttaa virheen pienentämiseksi oppimisen aikana ja painot päivitetään gradientin vastakkaiseen suuntaan. Gradientit voivat kasvaa tai pienentyä monien aika-askelten yli. Tällöin ne voivat joko kasvaa hyvin suuriksi (gradientin räjähtäminen) tai pienentyä lähes nolliin (gradientin katoaminen), mikä vaikeuttaa mallin oppimista [21]. Tämän vuoksi ne voivat aliarvioida aiempien syötteiden merkitystä ja jäädä jopa yksinkertaisempien mallien tasolle [10], [12]. Yleisimmin käytettyjä RNN-muunnoksia ovat LSTM-, GRU-verkot [6]. LSTM (Long Short-Term Memory) on yleisin RNN-muunnos ja se on kehitetty ratkaisemaan tavallisten RNN-verkkojen ongelmia, kuten pitkien riippuvuuksien heikkoa käsittelyä sekä gradientin katoamista ja räjähtämistä [12], [13]. LSTM:n muistisolu ja kolme porttia, unohtamis-, sisääntulo- ja ulostuloportti säätelevät, mitä tietoa säilytetään, päivitetään ja siirretään eteenpäin [6]. Tämän rakenteen ansiosta LSTM pystyy hyödyntämään pitkäkestoisia riippuvuuksia tehokkaammin kuin perinteiset RNN:t ja sitä käytetään esimerkiksi tekstiluokittelussa, jossa tarvitaan monipuolisen kielellisen

kontekstin huomioimista [6], [10]. GRU on LSTM:n kevyempi versio, joka käyttää kolmen portin sijasta kahta porttia, päivitysporttia ja nollausporttia. Päivitysportti päättää, kuinka paljon vanhaa tietoa pidetään mukana ja nollausportti määrää, mitä osia menneestä tiedosta unohdetaan. Näiden avulla muodostetaan ehdokastila, jota verrataan aiempaan tilaan. Lopullinen muistiarvo syntyy, kun päivitysportti yhdistää sopivassa suhteessa vanhan tilan ja uuden ehdokastilan [6]. Kuvassa 2.5 havainnollistetaan ketjumaista LSTM-mallia.

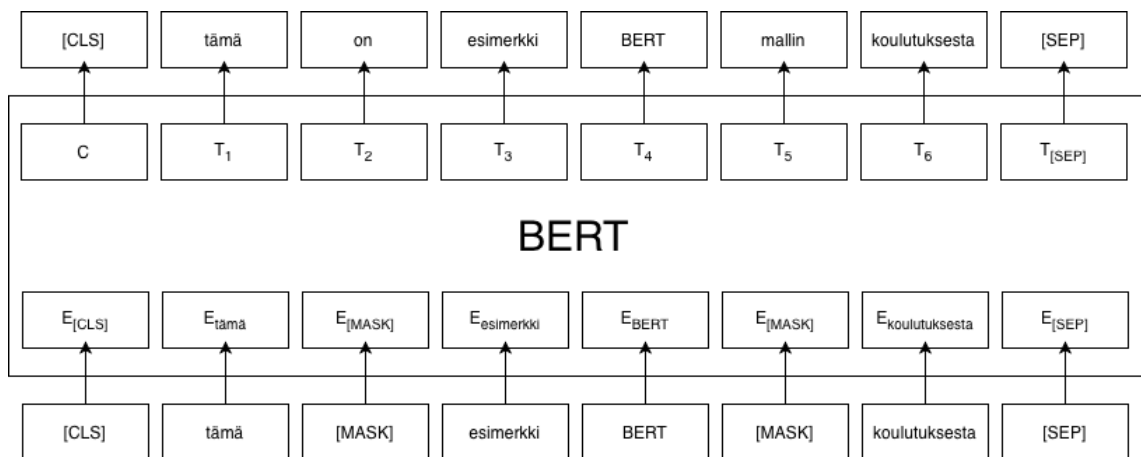


Kuva 2.5: Esimerkki ketjumaisesta LSTM-mallista, jossa jokainen solu vastaanottaa uuden syötteen sekä aiemman tilan ja säätelee, mitä tietoa säilytetään, päivitetään ja siirretään eteenpäin seuraavaan vaiheeseen. Malli mahdollistaa aiempien syötteiden vaikutuksen myöhempisiin ulostuloihin, mikä on keskeistä sarjallisen datan käsittelyssä. Lähde: Mukailtu [10]

Yksi RNN-verkkojen merkittävimmistä laskennallisista rajoitteista on tekstin sarjallinen käsittely, minkä vuoksi ne eivät kykene hyödyntämään tehokkaasti rinnakkaislaskentaa. Vaikka CNN-mallit ovat vähemmän sarjallisia, myös niiden laskennallinen kustannus kasvaa lauseen pituuden mukana, kun ne yrittävät mallintaa sanojen välisiä suhteita [10]. Transformer-arkkitehtuuri [34] ratkaisee nämä ongelmat käyttämällä itsehuomiointi (self-attention) -mekanismia. Siinä huomioarvot lasketaan samanaikaisesti kaikille lauseen sanoille matriisien avulla, mikä mahdollistaa pitkän kantaman riippuvuuksien tehokkaan mallintamisen. Tämän ansiosta Transformer-mallit skaalautuvat huomattavasti paremmin kuin RNN- tai CNN-pohjaiset mallit ja ne voidaan kouluttaa tehokkaasti suurilla tietomäärillä GPU- ja TPU-laitteistoilla [10]. Transformer-tyyppiset mallit ovat viime vuosina nous-

seet NLP:n valtavirtaan ja korvanneet RNN- ja LSTM-pohjaiset ratkaisut monissa tehtävissä, kuten luonnollisen kielen käsittelyssä [35], [36]. Syynä on erityisesti itsehuomiointi-mekanismien kyky huomioida etäiset sanasuhteet riippumatta siitä, esiintyvätkö ne tekstin alussa, keskellä tai lopussa [36]. Transformer-pohjaisten mallien suorituskky paranee merkittävästi niiden koon kasvaessa, mikä on ollut mahdollista kehittyneen laitteistotuen ansiosta [35]. Lisäksi läpimurtoja on saavutettu, kun malleja on alettu esikouluttaa valvomattomasti suurilla Internetistä kerätyillä tekstiaineistoilla, mikä on nostanut kielimallien suorituskvyn osittain ihmisen tasolle [10]. Transformer-pohjaiset esikoulutetut kielimallit ovat nousseet keskeiseksi tekniikaksi vuodesta 2018 lähtien. Ne edustavat muun muassa CNN- ja LSTM-pohjaisista malleista poikkeavaa arkkitehtuuria ja niissä hyödynnetään huomattavasti syvempiä verkkoja, sekä suurempia tekstiaineistoja esikoulutuksessa. Ensin mallit esikoulutetaan valvomattomasti oppimaan yleisiä kielen rakenteita, minkä jälkeen niitä hienosäädetään valvotusti eri tehtäviin. Tämä yhdistelmä on tuottanut selviä parannuksia monien NLP-tehtävien suorituskvyn [10]. Esikoulutetut mallit jaetaan yleensä autoregressiivisiin ja autoenkooderiperusteisiin malleihin [10]. Generative Pre-Training-malli (GPT) [37], jonka Radford et al. esittelivät vuonna 2018 oli yksi ensimmäisistä Transformer-arkkitehtuuriin perustuvista autoregressiivisistä generatiivisista malleista [36]. GPT hyödyntää Transformer-dekooderia, jossa self-attention toimii yksisuuntaisesti. Näin malli voi tuottaa tai generoida tekstiä, mikä sopii muun muassa kääntämiseen, tiivistämiseen ja kysymyksiin vastaamiseen [36]. GPT:n seuraajat, kuten GPT-3, kasvattavat alkuperäistä mallia huomattavasti suuremmilla parametrimäärillä. Kehittyneiden GPT-mallien kyky tuottaa pitkiä ja yhtenäisiä tekstejä on tehnyt siitä tärkeän kehitysaskeleen modernissa NLP:ssä [35]. BERT [38] on Googlen kehittämä kaksisuuntaiseen Transformer-arkkitehtuuriin perustuva kielimalli, jonka esittelivät Devlin et al. [10]. Toisin kuin autoregressiiviset mallit, kuten GPT, jotka ennustavat seuraavan sanan aiempien sanojen perusteella, BERT pystyy lukemaan

tekstin molempiin suuntiin ja huomioimaan koko lauseen kontekstin samanaikaisesti. Tämä auttaa sitä ymmärtämään sanojen merkitystä niiden koko kontekstissa [12]. BERT koulutetaan peitettyjen sanojen ennustamisella, jossa osa sanoista peitetään ja mallin on ennustettava ne. Lisäksi käytetään seuraavan lauseen ennustamista [10], [12], [36]. Kaksisuuntainen rakenne mahdollistaa suurempien aineistojen hyödyntämisen kuin perinteiset sarjalliset mallit [12]. Esikoulutuksen ansiosta BERT oppii kielen rakenteita tehokkaasti ja sitä voidaan myöhemmin hienosäätää eri tehtäviin ohjatulla oppimisella [12]. BERTistä on kehitetty myös parannettuja versioita kuten RoBERTa, joka on koulutettu suuremmalla aineistolla, sekä DistilBERT, joka on kevyempi ja nopeampi, mutta lähes yhtä suorituskykyinen [10]. Kuvassa 2.6 havainnollistetaan BERT-mallin kouluttamista.



Kuva 2.6: Esimerkki BERT-mallin kouluttamisesta. Kuvassa esitetään, miten syöte koostuu tokeneista (esim. [CLS], sanat ja [SEP]). Osa sanoista korvataan [MASK]-tokeneilla ja malli pyrkii ennustamaan nämä puuttuvat sanat kontekstin perusteella. Ylemmällä rivillä esitetään mallin tuottamat esitykset ja alemmalla rivillä alkupe-  
räinen syöte maskauksineen. Lähde: Mukailtu [38], [39]

# 3 Kirjallisuuskatsaus

## sentimenttianalyysin

### hyödyntämisestä tuotekehityksessä

#### 3.1 Koneoppimiseen perustuvan sentimenttianalyysin soveltaminen tuotekehityksessä

Asiakaspalautteen hyödyntäminen on tärkeä osa liiketoimintaa. Kuluttajien tuotearvostelut ja sosiaalisen median mielipiteet verkossa tarjoavat yrityksille konkreettista tietoa siitä, miten tyytyväisiä asiakkaat ovat ja miten hyvin tuotteet toimivat arjessa ja niiden avulla voidaan myös tunnistaa toistuvia ongelmia sekä kehityskohteita [32], [40], [41], [42], [43]. Tämän vuoksi sentimenttianalyysia ja koneoppimista hyödynnetään yhä enemmän tuotekehityksen tukena, sillä perinteiset menetelmät, kuten kyselyt tai fokusryhmät, ovat usein hitaita ja kalliita [32], [43], [44]. Tuotearvostelut ja sosiaalisen median sisällöt tarjoavatkin helposti saatavilla olevaa asiakaspalautetta, josta voidaan tunnistaa mielipiteitä tai kehityskohteita [14], [15], [32]. Lisäksi arvostelut ovat yleensä helpommin kerättäviä ja kustannustehokkaampia kuin kyselyt [44], [45]. Näin sentimenttianalyysia voidaan hyödyntää uusien tuotteiden kehityksen (NPD, New Product Development) tukena, koska sen avulla asiakasarvosteluista

ja muusta verkkoaineistosta on mahdollista tunnistaa asiakkaiden tarpeita ja mielipiteitä [15], [35].

Sentimenttianalyysi on tuotekehityksessä tärkeä menetelmä, koska se mahdollistaa asiakaspalautteen tulkinnan aspektitasolla [14], [15], [32]. Sen sijaan, että palaute tiivistettäisiin yhdeksi yleiseksi arvosteluksi, menetelmä tunnistaa tunnesävyyn tekstistä, jossa käsitellään tuotteen yksittäisiä ominaisuuksia, kuten esimerkiksi älypuhelimien kameraa [32], sen näyttöä [32] tai tietokoneen hintaa [15]. Tämän ansiosta voidaan tunnistaa täsmällisesti, mitkä ominaisuudet koetaan vahvuuksiksi ja mitkä aiheuttavat tyytymättömyyttä [15], [42]. Aspektitason analyysi tukee kohdennettujen parannustoimenpiteiden suunnittelua sekä kehitystyön priorisointia ja päätöksentekoa [14], [15], [32], [42].

Koneoppimiseen perustuvaa sentimenttianalyysia hyödynnetään tuotekehityksessä erityisesti siksi, että sen avulla suuri määrä asiakaspalautetta voidaan muuttaa käytännön kehitystoimenpiteiksi. Kirjallisuudessa menetelmiä käytetään asiakastarpeiden ja vaatimusten tunnistamiseen sekä käyttäjätyytyväisyyden ja mieltymysten seuraamiseen, mikä tukee tuotelaadun parantamista [15], [32], [40], [41], [44]. Osa tutkimuksesta korostaa myös innovaatiohyötyjä, jossa mielipiteitä analysoimalla voidaan tunnistaa uusia käyttötapauksia ja tuoteideoita [43]. Tuoteominaisuuksia voidaan myös luokitella ja priorisoida Kano-mallin avulla [32], [45] ja selitettävä tekoäly tukee [45]. Kano-malli ei itsessään ole sentimenttianalyysimenetelmä, mutta sitä voidaan yhdistää sentimenttianalyysiin. Lisäksi menetelmiä voidaan soveltaa asiakassegmentointiin, jolloin kehitystä voidaan kohdentaa eri käyttäjäryhmille [15]. Nykyaikaiset transformer-pohjaiset kielimallit laajentavat käyttökohdetta entisestään tukemalla esimerkiksi tiivistämistä, oivallusten poimintaa ja ideointia innovaatio- tai tuotekehitysprosessissa [35].

Tutkimuskirjallisuudessa sentimenttianalyysia hyödynnetään tuotekehityksen tukena erityisesti asiakastiedon tulkintaan verkkokaupan tuotearvioista [40], [41],

[42], sosiaalisesta mediasta [15], [43] sekä asiakasvideoista [44]. Monissa töissä korostuu, ettei pelkkä sentimenttiluokittelu usein riitä, vaan tuotekehitystä varten on ensin tunnistettava, mistä ominaisuuksista tai teemoista palautteessa puhutaan. Tähän käytetään yleisesti aihemallinnus-menetelmiä, joissa LDA nousee toistuvasti keskeiseksi ratkaisuksi. Tämän avulla arviot voidaan jakaa tuoteominaisuuksiksi tai teemoihin [14], [32], [44]. Alan tutkimukset kattavat menetelmiä perinteisistä koneoppimismalleista syväoppimiseen ja transformer-pohjaisiin malleihin. Perinteisistä luokittelijoista käytetään muun muassa tukivektorikonetta (SVM) [41], [42], logistista regressiota (LR) [40], [41], päätöspuita (DT) [40], [41], satunnaismetsiä (RF) [40], [41] sekä Naiivia Bayesia (NB) [42] ja sen variantteja (MNB) [40]. Kun tavoitteena on jäljitellä aineiston monimutkaisia riippuvuuksia ja kontekstia, siirrytään neuroverkkoihin ja syväoppimismalleihin, kuten CNN-malleihin [32], [40], [41], [42], LSTM- [41], [42], [44] ja Bi-LSTM-malleihin [40], [42], [44] sekä GRU-malleihin [41]. Viimeaikaisissa tutkimuksissa korostuvat lisäksi transformer-pohjaiset mallit, joissa esimerkiksi BERT [15], [40], [41], XLNet [40], [41] ja RoBERTa [41] esitetään vaihtoehtoiksi. Tuotekehityskäytössä ilmenee myös mallien selitettävyyttä, jota voidaan tukea esimerkiksi LIME (Local Interpretable Model-Agnostic Explanations) [40], [41] tai SHAP (Shapley Additive Explanation) -menetelmillä [45]. Neuroverkot soveltuvat hyvin monimutkaisten ja epälineaaristen suhteiden oppimiseen, mutta mallien sisäinen toiminta jää tyypillisesti vaikeasti tulkittavaksi, sillä ne toimivat luonteeltaan "black box" -malleina [45]. Tämän rajoitteen vuoksi selitettävän tekoälyn menetelmiä, kuten LIME hyödynnetään mallien toiminnan ymmärrettävyyden parantamiseksi. Näiden menetelmien avulla voidaan tarkastella, mitkä muutokset vaikuttavat mallin antamiin tuloksiin tietyissä tilanteissa. Selitettävä tekoäly on erityisen tärkeä sovelluksissa, joissa mallin tuottamien tulosten perusteltavuus ja läpinäkyvyys ovat vähintään yhtä tärkeitä kuin ennustetarkkuus [41].

## 3.2 Käytetyt aineistot ja niiden vaikutukset menetelmiin

Sentimenttianalyysi tuotekehityksen tukena perustuu vahvasti käyttäjien luomaan sisältöön (UGC, User-Generated Content), jota kertyy runsaasti erilaisilta verkkoalustoilta, kuten blogeista, sosiaalisesta mediasta ja verkkokauppasivustoista [40], [41]. Asiakkaat jakavat kokemuksiaan ja odotuksiaan yhä aktiivisemmin julkisilla alustoilla, minkä vuoksi tutkimuksissa keskeisiä tietolähteitä ovat verkkokauppojen tuotearvostelut ja luokitukset, kuten Amazon [32], [40], [41] ja Tmall [32], sekä sosiaalisen median alustat, kuten X (ent. Twitter) [43], YouTube ja TikTok [44] sekä erilaiset foorumit [40], [41]. Erityisesti sosiaalinen media voi tarjota dynaamista tai uutta tietoa asiakaspalautteista [43]. Verkkokauppa-arvosteluissa on usein analyysia helpottavia rakenteita, kuten tähtiluokitukset [32], [41] tai vahvistetun oston merkinnät [32]. Lisäksi käyttäjien luomat videot tuovat analyysiin useita datamuotoja, kuten tekstiä, ääntä ja kuvia, mikä edellyttää kehittyneempiä analyysimenetelmiä sentimenttianalyysin ohella [44].

Sentimenttianalyysissa menetelmän valinta riippuu aineiston luonteesta ja analyysitehtävän vaatimuksista [42]. Koska aineisto on usein jäsentämätöntä tekstiä, se vaatii tyypillisesti esikäsittelyä, kuten tokenisointia tai epäolennaisten merkkien poistamista [41], [42]. Tämän jälkeen tarvitaan piirteiden poimintaa, jossa data muunnetaan piirrevektoreiksi esimerkiksi BOW [40], [43], TF-IDF [40], [41], [42], [43] tai sanaupotusten, kuten Word2Vec [41], [42], [43], GloVe [43] tai FastText [41] avulla. Suurissa aineistoissa voidaan hyödyntää syväoppimis- ja transformer-pohjaisia malleja niiden kyvyn vuoksi käsitellä laajoja tietomääriä [15], [35]. Useita datamuotoja sisältävässä datassa, kuten videoissa, on usein hyödyllistä käyttää menetelmiä, jotka kykenevät yhdistämään eri lähteistä saatavaa tietoa sentimentin tunnistamisen tueksi, jolloin analyysi ei perustu pelkästään tekstiin [44]. Tuloksia

voidaan parantaa lisäämällä mukaan tuotetietoja tai -nimiä, jotka auttavat mallia ymmärtämään aineistojen kontekstia [41]. Lisäksi asiakasvaatimusten tunnistamista voidaan tehostaa käyttämällä selkeää luokittelua, jossa huomioidaan myös neutraali sentimentti [32]. Tulosten tulkinnassa on myös tärkeä huomioida, että sentimenttiai-  
neisto voi olla epätasapainoista, tyypillisesti jonkun tietyn sentimentin painottuessa [40], [43]. Epätasapainoinen aineisto voi aiheuttaa vinoumia datan jakaumassa [42], mikä voi johtaa siihen, että mallit suosivat enemmistöluokkaa ja antavat optimistisen kuvan suorituskyvystä.

### 3.3 Keskeiset tulokset ja vertailu

Sentimenttianalyysin koneoppimismenetelmien suorituskyky vaihtelee sen mukaan, kuinka hyvin ne pystyvät mallintamaan luonnollisen kielen monimutkaisia riippuvuuksia [40], [41], [42], [44]. Menetelmiä arvioidaan tyypillisesti mittareilla kuten accuracy, F1-score, precision ja recall [40]. Tulosten perusteella mikään yksittäinen lähestymistapa ei osoittaudu parhaaksi kaikissa tarkastelluissa asetelmissä, vaan onnistuminen riippuu vahvasti käytetystä aineistosta, luokittelutarkkuudesta (esim. 5-luokkaisen asteikon yhdistäminen 3-luokkaiseksi), sekä piirteiden esitystavasta [40], [41].

Perinteiset koneoppimismenetelmät, kuten LR, SVM tai RF ovat usein laskennallisesti kevyitä ja soveltuvat hyvin skaalautuviin sovelluksiin verrattuna esimerkiksi transformer-pohjaisiin menetelmiin [41]. Lisäksi osa malleista tarjoaa jonkinlaista tulkittavuutta, kuten esimerkiksi logistisen regression ennusteet [40], [41]. Heikkoutena on kuitenkin rajallinen kyky huomioida konteksti, mikä voi heikentää suorituskykyä erityisesti moniluokkaisissa sentimenttianalyysitehtävissä [40]. Tätä voi parantaa luokkien vähentäminen. Tämä näkyy esimerkiksi siinä, että viiden luokan luokittelussa SVM saavutti FastText-upotuksella mobiili elektroniikan aineistossa

noin 84 % tarkkuuden [41], kun taas luokkien vähentäminen kolmeen luokkaan paransi tulosta selvästi ja nosti tarkkuuden 92 % [41].

Syväoppimismallit, kuten CNN ja Bi-LSTM voivat oppia tekstistä automaattisesti piirteitä ja mallintaa tekstin sarjallisia riippuvuuksia, mikä on olennaista arvostelujen kontekstin ymmärtämisessä [40]. Tämä ei kuitenkaan aina pidä paikkansa, sillä esimerkiksi CNN-LSTM saavutti 92 % tarkkuuden kolmen luokan sentimenttluokittelussa FastText-upotuksella mobiili elektroniikan aineistossa, mikä on sama taso kuin perinteisellä SVM-mallilla [41]. Syväoppimismallit tarvitsevat usein merkityn opetusdatan [32], [44], eikä kaikkien syväoppimismallien laskennallinen kuormitus eroa merkittävästi perinteisistä menetelmistä [41]. Sen sijaan syväoppimismallien tulosten tulkittavuus voi olla haastavampaa [45].

Transformer-pohjaiset mallit, kuten BERT, RoBERTa tai XLNet ovat usein vahvoja menetelmiä monissa tutkimuksissa, sillä ne pystyvät huomaamaan tekstin viivahteita ja suoriutuvat monista eri NLP-tehtävistä [41]. Ali et al. [40] tutkimuksessa BERT-malli saavutti esimerkiksi Amazon-arvostelujen luokittelussa parhaimman tarkkuuden, päihittäen useita perinteisiä ja syväoppivia malleja. Toisaalta näiden mallien suorituskyky riippuu aineistosta ja vertailuasetelmasta. Hashmi et al. [41] tutkimuksessa RoBERTa saavutti 90 % tarkkuuden kolmen luokan sentimenttluokittelussa mobiili elektroniikan aineistossa, mikä on hieman heikompi kuin FastText-upotuksesta hyödyntäneellä SVM-mallilla. Näiden mallien keskeiset haasteet liittyvät kuitenkin laskennalliseen kustannukseen ja koulutuksen hitauteen [41], mahdolliseen ylisovittamiseen [41], sekä mallien tulosten tulkittavuuteen [40], [45]. Lisäksi erityisesti generatiivisiin malleihin, kuten GPT liittyy riski tuottaa vakuuttavaa, mutta epäluotettavaa sisältöä, mikä korostaa asiantuntijavalvonnan merkitystä [35].

Tarkastelluissa tutkimuksissa yhdistelmämallit (ensemble) ja optimointipohjaiset koneoppimismenetelmät saavuttivat parhaat tulokset, kun ne yhdistivät eri lähestymistapojen vahvuuksia. Esimerkiksi FastXCatStack, eli FastText, XGBoost

ja CatBoost -menetelmien yhdistelmä saavutti jopa 93 % tarkkuuden kolmen luokan sentimenttiluokittelussa mobiili elektroniikan aineistossa ja oli parhain kolmessa neljästä arvioidusta aineistosta [41]. Vastaavasti Rasappan et al. [42] tutkimuksessa optimointialgoritmeilla hienosäädetty EGJO-LSTM (Enhanced Golden Jackal Optimizer-pohjainen LSTM) raportoi parhaimman tarkkuuden kaikista tutkimuksen menetelmistä tietyssä viiden luokan luokittelutehtävässä. Tämä korostaa piirteiden esitystavan käytännön merkitystä [41], [42].

## 4 Pohdinta

Koneoppimiseen perustuva sentimenttianalyysi tarjoaa tuotekehitykselle keinon tunnistaa suurista asiakaspalauteaineistoista mielipiteitä, ongelma- ja kehityskohteita kustannustehokkaasti. Erityisesti aspektitason analyysi mahdollistaa havaintojen kohdentamisen tuoteominaisuuksiin ja niitä koskeviin päätöksiin. Samalla tämän työn kirjallisuuskatsaus vahvistaa, että käytännön hyödyn saavuttaminen on harvoin kiinni yhdestä parhaasta mallista, vaan kokonaisuudesta, jossa itse mallinustehtävä, data, piirre-esitys, esikäsittely ja käyttökonteksti määrittävät lopputuloksen.

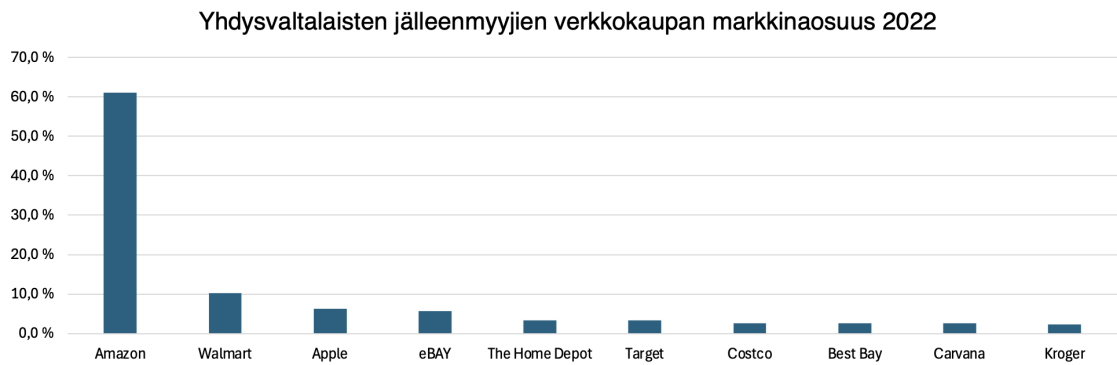
Yksi keskeinen haaste on datan soveltuvuus käyttötarkoitukseensa sekä tulosten yleistettävyyden. Tutkimuskirjallisuudessa korostuvat tietyt helposti saatavilla olevat UGC-lähteet, erityisesti verkkokauppojen tuotearvostelut [32], [40], [41], sekä sosiaalisen median kanavat [43] ja osassa töistä hyödynnetään myös videoalustoja [44]. Tämä on ymmärrettävää, koska lähteet tarjoavat suuria aineistoja ja verkkokaupoissa on usein analyysia helpottavia rakenteita [32], [41]. Samalla tällainen painottuminen voi ohjata menetelmävertailuja kohti yksinkertaisempia asetelmia ja jättää vähemmälle huomiolle tilanteet, joissa palaute on lyhyttä, kontekstiltään puutteellista tai peräisin useista eri lähteistä. Lisäksi kielten sekä tuote- ja käyttäjäryhmien väliset erot voivat olla merkittäviä, minkä vuoksi malli, joka toimii erinomaisesti yhdessä verkkokauppakategoriassa, ei välttämättä yleisty suoraan toiseen ilman uudelleen koulutusta tai hienosäätöä. Tämän tilanteen taustalla on yleisempi koneoppi-

misen haaste, jossa menetelmät ovat usein aihealueeseen sidonnaisia ja suorituskyky heikkenee, kun hyvin yhdellä aineistolla toimivaa menetelmää sovelletaan erilaiseen aineistoon [12].

Muita käytännön haasteita ovat merkityn opetusdatan saatavuus sekä arviointien vertailtavuus. Ohjatut koneoppimismenetelmät edellyttävät tyypillisesti suuria määriä merkittävää dataa ja tämän rajallinen saatavuus on keskeinen haaste etenkin uusissa aihealueissa. Lisäksi kirjallisuudessa tuloksia raportoidaan useilla eri mittareilla sekä vaihtelevilla luokitteluasetelmilla, kuten viisi- ja kolmiluokkaisilla luokituksilla [40], [41]. Tämä asettaa haasteita mallien tulosten suoraan vertailuun eri tutkimusten välillä. Tästä seuraa tuotekehityksen näkökulmasta tärkeä huomio, jossa arviointiasetelman soveltuvuus käytännön tarpeisiin nousee esiin. Jos arviointiasetelma ei esimerkiksi huomioi neutraalin sentimentin merkitystä asiakasvaatimusten tunnistamisessa, tämä voi johtaa sentimenttitulkinnan vääristymiseen ja arvokkaan tiedon menetykseen päätöksenteossa [32].

Tutkimusten aineistoissa toistui selkeä kaava, jossa englanninkielisten tuotearvostelujen lähteenä käytettiin Amazonin tuotearvosteluja. Tätä voidaan selittää Amazonin vahvalla asemalla Yhdysvaltojen verkkokaupassa, jota Hashmi et al. [41] havainnollistavat esittämällä vuoden 2022 markkinaosuusvertailun suurimmista yhdysvaltalaisista verkkokauppatoimijoista. Tämä tukee oletusta siitä, että Amazonissa asiakaspalautetta kertyy poikkeuksellisen paljon eri tuoteryhmistä. Tällainen laajuus on koneoppimisen näkökulmasta houkutteleva, koska se mahdollistaa suurten koulutusaineistojen rakentamisen ja menetelmien vertailun vertailukelpoisissa asetelmissa. Lisäksi Amazon-aineistojen käyttö on käytännössä vaivatonta, sillä niitä on jaettu valmiina kokonaisuuksina esimerkiksi Kagglissa, mikä helpottaa niiden käyttöönottoa ja parantaa tutkimusten vertailtavuutta. Samalla tämä voi kuitenkin johtaa tutkimuksissa käytettyjen aineistojen yksipuolistumiseen, jossa menetelmät kehitetään erityisesti Amazonille ominaiseen tekstiin ja asetelmaan, mikä voi heiken-

tää tulosten yleistettävyyttä muihin alustoihin ja asetelmiin. Kuvassa 4.1 havainnollistetaan yhdysvaltalaisien jälleenmyyjien verkkokaupan markkinaosuutta vuodelta 2022.



Kuva 4.1: Kuvassa havainnollistetaan yhdysvaltalaisien jälleenmyyjien verkkokaupan markkinaosuutta 2022. Nimi "Best Bay" on esitetty sellaisenaan kuin se esiintyi alkuperäislähteessä, eikä sen oikeellisuuteen oteta tässä kantaa. Lähde: Mukailtu [41]

Menetelmävalinnoissa tutkimukset tukevat tyypillistä tasapainoa, jossa perinteiset mallit ovat usein kevyitä ja helpommin selitettäviä, kun taas syväoppimiseen ja transformer-pohjaisiin malleihin perustuvat lähestymistavat käsittelevät paremmin kontekstia ja monimutkaisia riippuvuuksia. Vastapainona ne vaativat enemmän dataa ja laskentatehoa ja niiden tulokset ovat vaikeammin selitettävissä. Lisäksi katsauksen perusteella esikäsittely ja piirre-esitys eivät ole ainoastaan teknisesti välttämättömiä vaiheita, vaan usein ratkaisevia tekijöitä lopputuloksen tarkkuudessa. Optimaalinen lähestymistapa tässä vaikuttaa vaihtelevan käytetyn mallin ja aineiston mukaan ja käytännössä sopivat ratkaisut selviävät usein kokeilujen ja vertailun kautta. Tämän vuoksi yksittäisen tutkimusartikkelin parhaiden tulosten siirtäminen ja toistaminen sellaisenaan organisaation omassa datassa voi osoittautua hankalaksi.

Tuotekehityksen kannalta erityisen merkittävä haaste on tulkittavuuden puute, sillä pelkkä korkea ennustetarkkuus ei käytännössä riitä, jos tuloksia ei voida muuntaa konkreettisiksi toimenpiteiksi. Tässä aspektitason analyysi ja aihehallin-  
nusmenetelmät, kuten LDA, auttavat hahmottamaan mistä ominaisuuksista ja tee-

moista käyttäjät puhuvat, kun taas selitettävän tekoälyn menetelmät, kuten LIME ja SHAP, voivat auttaa ymmärtämään, miksi malli tuottaa tietyn sentimenttiluokituksen. Käytännössä haaste on kuitenkin se, että selitys ei automaattisesti tarkoita päätöksenteon tukea, vaan organisaation on kyettävä muun muassa priorisoimaan löydökset sen mukaan, mitkä ovat kriittisiä ja mitkä toissijaisia. Tässä voi toimia apuna esimerkiksi Kano-malli, mutta lisäksi haasteena on tulosten luotettavuuden varmistaminen, jotta ne eivät heijastaisi vain äänekkäimpien käyttäjien mielipiteitä tai aineiston vinoumia, kuten positiivisten arvostelujen ylikorostumista.

## 5 Yhteenveto

Tässä kandidaatintutkielmassa tarkasteltiin, miten koneoppimiseen perustuvaa sentimenttianalyysia hyödynnetään tuotteiden kehittämisen tukena. Tutkielman lähtökohtana on, että asiakkaiden mielipiteet tuotteista ovat keskeinen osa tuotekehitystä ja verkossa kertyvä käyttäjien tuottama sisältö, kuten tuotearvostelut ja sosiaalisen median julkaisut tarjoavat runsaasti tietoa asiakkaiden kokemuksista ja odotuksista. Koneoppimisen avulla palauteaineistoja voidaan käsitellä tehokkaasti ja erityisesti aspekti- tai ominaisuustason sentimenttianalyysi mahdollistaa havaintojen kohdentamisen suoraan tuotteiden yksittäisiin ominaisuuksiin. Seuraavaksi esitetään vastaukset tutkimuksen alussa asetettuihin tutkimuskysymyksiin.

**Tutkimuskysymys 1:** Kirjallisuuskatsauksen perusteella sentimenttianalyysissa tuotekehityksen tukena hyödynnetään monipuolisesti koneoppimismenetelmiä perinteisistä luokittelijoista syväoppimiseen ja transformer-pohjaisiin malleihin. Käytettyjä perinteisiä malleja ovat muun muassa SVM, LR, DT, RF sekä NB/MNB, kun taas syväoppimismenetelmiin kuuluvat CNN-, LSTM-, Bi-LSTM- ja GRU-mallit. Viimeaikaisissa tutkimuksissa hyödynnetään erityisesti transformer-pohjaisia malleja, kuten BERT, XLNet ja RoBERTa ja tekstiaineistojen aiheiden tunnistamista voidaan tukea LDA-menetelmän avulla. Näitä menetelmiä sovelletaan tuotekehityksessä erityisesti asiakastarpeiden ja -mieltymysten tunnistamiseen, uusien tuotteiden löytämiseen, käyttäjätyytyväisyyden seurantaan sekä tuuteominaisuuksien arviointiin aspektitason sentimenttianalyysin avulla. Lisäksi sentimenttianalyysi-

sia hyödynnetään tuoteominaisuuksien priorisoinnissa, asiakassegmentoinnissa ja innovaatioiden tukena.

**Tutkimuskysymys 2:** Tutkimuksissa keskeisiksi datan lähteiksi nousivat verkkokauppojen tuotearvostelut ja -luokitukset, kuten Amazon tai Tmall, sekä sosiaalisen median alustat, kuten X ja osassa tutkimuksista myös videoalustojen käyttäjäpalaute, kuten YouTube tai TikTok. Aineiston ominaisuudet ohjaavat merkittävästi menetelmävalintoja. Suurissa ja osittain rakenteisissa tekstiaineistoissa voidaan hyödyntää sekä perinteisiä koneoppimismenetelmiä, että syväoppimis- ja transformer-pohjaisia malleja, kun taas pienemmissä tai kohinaisemmissa aineistoissa nousevat esiin huolellinen esikäsittely ja piirre-esitysten merkitys. Tyypillisiä piirre-esityksiä ovat BOW- ja TF-IDF-menetelmät sekä sanaopetusmenetelmät, kuten Word2Vec, GloVe ja FastText. Syväoppivat ja transformer-pohjaiset menetelmät soveltuvat suuriin ja kontekstia sisältäviin aineistoihin. Useita datamuotoja sisältävä palaute edellyttää kehittyneempiä menetelmiä ja lisäksi aineistojen luokitteluasetelma, sekä luokkien epätasapaino vaikuttavat käytettäviin malleihin ja arviointiin.

**Tutkimuskysymys 3:** Tutkimuskirjallisuuden perusteella eri sentimenttianaalysin koneoppimismenetelmien vahvuudet ja heikkoudet tuotteiden kehittämisen tukemisessa liittyvät niiden kykyyn mallintaa kielen kontekstia suhteessa laskennallisiin kustannuksiin ja tulosten tulkittavuuteen. Mikään yksittäinen menetelmä ei ole kaikissa asetelmissa paras, vaan suorituskyky riippuu aineistosta, tehtävän määrittelystä ja piirre-esityksistä. Perinteiset koneoppimismenetelmät, kuten logistinen regressio ja SVM, ovat laskennallisesti kevyitä, skaalautuvia ja usein tulkittavia, mikä tukee tuotekehityksen päätöksentekoa, mutta niiden kyky huomioida kielellinen konteksti on rajallinen. Syväoppimismallit, kuten CNN- ja Bi-LSTM-arkkitehtuurit, mallintavat tekstin rakenteita ja riippuvuuksia paremmin, mutta edellyttävät enemmän koulutusdataa ja laskentatehoa ja ovat usein heikommin tulkittavia. Transformer-pohjaiset mallit, kuten BERT, RoBERTa ja XLNet, tarjoavat

tyypillisesti parhaan suorituskyvyn kielellisesti vaativissa asetelmissa, mutta niiden käyttöä rajoittavat korkeat laskennalliset kustannukset, mahdollinen ylisovittaminen ja tulosten heikko läpinäkyvyys. Yhdistelmämallit ja optimointiin perustuvat lähestymistavat voivat saavuttaa vielä paremman suorituskyvyn yhdistämällä eri menetelmien vahvuuksia ja hyödyntämällä piirre-esityksiä tehokkaasti.

Työn rajoitteet liittyvät ennen kaikkea kirjallisuuskatsauksen lähtökohtiin ja rajauksiin. Katsaus koottiin valituista tietokannoista ja rajattiin pääosin vuosille 2021–2025 ajantasaisuuden varmistamiseksi. Rajausta parantaa katsauksen tuoreutta, mutta voi samalla sulkea pois joitakin aiempia, edelleen relevantteja tutkimuksia ja vertailuja. Lisäksi tutkimusten vaihtelevuus aineistojen, asetelmien, esikäsittelyn ja mittareiden osalta rajoittaa mahdollisuuksia tehdä täysin yksiselitteisiä johtopäätöksiä siitä, että jokin tietty malli olisi selvästi paras. Tästä huolimatta kokonaiskuva on selkeä, sillä tuotekehityksessä onnistunut sentimenttianalyysi perustuu prosessiin, jossa tekninen suorituskyky, datan laatu, tulkittavuus ja päätöksentekoon kytkeytyminen ovat yhtä keskeisiä.

Jatkotutkimuksen kannalta keskeinen suunta on laajeneminen kohti monipuolisempia aineistoja ja suurempia kielimalleja. Käyttäjien tuottama sisältö ei ole vain tekstiä, vaan myös esimerkiksi videoita, joissa palaute ilmenee tekstin lisäksi äänenä ja kuvana. Tällöin analyysi ei ole enää pelkästään sentimenttianalyysia, vaan edellyttää useiden eri mallien samanaikaista hyödyntämistä [44]. Tämä kasvattaa menetelmien ja arvioinnin vaatimuksia ja korostaa eri datalähteitä yhdistävien mallien tarvetta. Lisäksi transformer-pohjaisten mallien hyödyntäminen kohdennetuissa sovelluksissa todennäköisesti yleistyy, kun malleja hienosäädetään tiettyyn toimialaan ja käyttötarkoitukseen. Mallien täydentäminen kontekstilla, kuten tuotteeseen tai palveluun liittyvällä taustatiedolla, voi parantaa analyysin arvoa päätöksenteon tukena verrattuna pelkkään yleiseen sentimenttiluokitteluun.

# Lähdeluettelo

- [1] L. Ashbaugh ja Y. Zhang, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning", *Computers*, vol. 13, nro 12, 2024. DOI: <https://doi.org/10.3390/computers13120340>.
- [2] K. Schouten ja F. Frasincar, "Survey on Aspect-Level Sentiment Analysis", *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, nro 3, s. 813–830, 2016. DOI: <https://doi.org/10.1109/TKDE.2015.2485209>.
- [3] K. Ravi ja V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications", *Knowledge-Based Systems*, vol. 89, s. 14–46, 2015. DOI: <https://doi.org/10.1016/j.knosys.2015.06.015>.
- [4] L. Zhang, S. Wang ja B. Liu, "Deep learning for sentiment analysis: A survey", *WIREs Data Mining and Knowledge Discovery*, vol. 8, nro 4, e1253, 2018. DOI: <https://doi.org/10.1002/widm.1253>.
- [5] B. Ozyurt ja M. A. Akcayol, "A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA", *Expert Systems with Applications*, vol. 168, s. 114231, 2021. DOI: <https://doi.org/10.1016/j.eswa.2020.114231>.
- [6] A. Yadav ja D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review", *Artificial Intelligence Review*, vol. 53, nro 6, s. 4335–4385, 2020. DOI: <https://doi.org/10.1007/s10462-019-09794-5>.

- [7] ”Mining customer requirements from online reviews: A product improvement perspective”, *Information Management*, vol. 53, nro 8, s. 951–963, 2016. DOI: <https://doi.org/10.1016/j.im.2016.06.002>.
- [8] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu ja A. Abusorrah, ”Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods”, *IEEE Transactions on Computational Social Systems*, vol. 7, nro 6, s. 1358–1375, 2020. DOI: <https://doi.org/10.1109/TCSS.2020.3033302>.
- [9] A. Abdi, S. M. Shamsuddin, S. Hasan ja J. Piran, ”Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion”, *Information Processing Management*, vol. 56, nro 4, s. 1245–1259, 2019. DOI: <https://doi.org/10.1016/j.ipm.2019.02.018>.
- [10] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu ja J. Gao, ”Deep Learning-based Text Classification: A Comprehensive Review”, *ACM Comput. Surv.*, vol. 54, nro 3, 2021. DOI: <https://doi.org/10.1145/3439726>.
- [11] E. Cambria, B. Schuller, Y. Xia ja C. Havasi, ”New Avenues in Opinion Mining and Sentiment Analysis”, *IEEE Intelligent Systems*, vol. 28, nro 2, s. 15–21, 2013. DOI: <https://doi.org/10.1109/MIS.2013.30>.
- [12] N. Braig, A. Benz, S. Voth, J. Breitenbach ja R. Buettner, ”Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data”, *IEEE Access*, vol. 11, s. 14 778–14 803, 2023. DOI: <https://doi.org/10.1109/ACCESS.2023.3242234>.
- [13] N. C. Dang, M. N. Moreno-García ja F. De la Prieta, ”Sentiment Analysis Based on Deep Learning: A Comparative Study”, *Electronics*, vol. 9, nro 3, 2020. DOI: <https://doi.org/10.3390/electronics9030483>.

- [14] S. Güneş, "Extracting design knowledge from online product reviews to support design creativity", *International Journal of Design Creativity and Innovation*, vol. 11, nro 4, s. 273–293, 2023. DOI: <https://doi.org/10.1080/21650349.2023.2260826>.
- [15] M. Maghsoudi, N. Mohammadi ja M. Bakhtiari, "A novel approach to customer segmentation for product development on social media data: integrating aspect-based sentiment analysis and text mining", *Knowledge-Based Systems*, vol. 328, s. 114269, 2025. DOI: <https://doi.org/10.1016/j.knosys.2025.114269>.
- [16] W. Medhat, A. Hassan ja H. Korashy, "Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal*, vol. 5, nro 4, s. 1093–1113, 2014. DOI: <https://doi.org/10.1016/j.asej.2014.04.011>.
- [17] A. Esuli ja F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining", teoksessa *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, N. Calzolari et al., toim., 2006. url: <https://aclanthology.org/L06-1225/>.
- [18] G. A. Miller, "WordNet: a lexical database for English", *Commun. ACM*, vol. 38, nro 11, s. 39–41, 1995. DOI: <https://doi.org/10.1145/219717.219748>.
- [19] M. Taboada, J. Brooke, M. Tofiloski, K. Voll ja M. Stede, "Lexicon-Based Methods for Sentiment Analysis", *Computational Linguistics*, vol. 37, nro 2, s. 267–307, 2011. DOI: [https://doi.org/10.1162/COLI\\_a\\_00049](https://doi.org/10.1162/COLI_a_00049).
- [20] E. Cambria, "Affective Computing and Sentiment Analysis", *IEEE Intelligent Systems*, vol. 31, nro 2, s. 102–107, 2016. DOI: <https://doi.org/10.1109/MIS.2016.31>.

- [21] Y. LeCun, Y. Bengio ja G. Hinton, "Deep Learning", *Nature*, vol. 521, nro 7553, s. 436–444, 2015. DOI: <https://doi.org/10.1038/nature14539>.
- [22] A. Tripathy, A. Agrawal ja S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach", *Expert Systems with Applications*, vol. 57, s. 117–126, 2016. DOI: <https://doi.org/10.1016/j.eswa.2016.03.028>.
- [23] Y. Zhang, R. Jin ja Z.-H. Zhou, "Understanding bag-of-words model: a statistical framework", *International Journal of Machine Learning and Cybernetics*, vol. 1, s. 43–52, 2010. DOI: <https://doi.org/10.1007/s13042-010-0001-0>.
- [24] D. M. El-Din, "Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis", *International Journal of Advanced Computer Science and Applications*, vol. 7, nro 1, 2016. DOI: <https://doi.org/10.14569/IJACSA.2016.070134>.
- [25] D. M. Blei, A. Y. Ng ja M. I. Jordan, "Latent dirichlet allocation", *The Journal of Machine Learning Research*, vol. 3, s. 993–1022, 2003. url: <https://dl.acm.org/doi/10.5555/944919.944937>.
- [26] T. Mikolov, K. Chen, G. Corrado ja J. Dean, "Efficient Estimation of Word Representations in Vector Space", 2013. DOI: <https://doi.org/10.48550/arXiv.1301.3781>.
- [27] J. Pennington, R. Socher ja C. Manning, "GloVe: Global Vectors for Word Representation", teoksessa *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang ja W. Daelemans, toim., 2014, s. 1532–1543. DOI: <https://doi.org/10.3115/v1/D14-1162>.
- [28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever ja R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting", *J. Mach.*

- Learn. Res.*, vol. 15, nro 1, s. 1929–1958, 2014. url: <https://dl.acm.org/doi/10.5555/2627435.2670313>.
- [29] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed ja J. A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on  $\chi^2$  Statistical Model and Optimally Configured Deep Neural Network", *IEEE Access*, vol. 7, s. 34938–34945, 2019. DOI: <https://doi.org/10.1109/ACCESS.2019.2904800>.
- [30] B. Pang, L. Lee ja S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", 2002. DOI: <https://doi.org/10.48550/arXiv.cs/0205070>.
- [31] R. McDonald, K. Hannan, T. Neylon, M. Wells ja J. Reynar, "Structured Models for Fine-to-Coarse Sentiment Analysis", teoksessa *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, A. Zaenen ja A. van den Bosch, toim., 2007, s. 432–439. url: <https://aclanthology.org/P07-1055/>.
- [32] M. Zhang, L. Sun, G. A. Wang, Y. Li ja S. He, "Using neutral sentiment reviews to improve customer requirement identification and product design strategies", *International Journal of Production Economics*, vol. 254, s. 108641, 2022. DOI: <https://doi.org/10.1016/j.ijpe.2022.108641>.
- [33] T. A. Rana ja Y.-N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey", *Artificial Intelligence Review*, vol. 46, nro 4, s. 459–483, 2016. DOI: <https://doi.org/10.1007/s10462-016-9472-z>.
- [34] A. Vaswani et al., "Attention Is All You Need", 2017. DOI: <https://doi.org/10.48550/arXiv.1706.03762>.
- [35] S. G. Bouschery, V. Blazevic ja F. T. Piller, "Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language mo-

- dels”, *Journal of Product Innovation Management*, vol. 40, nro 2, s. 139–153, 2023. DOI: <https://doi.org/10.1111/jpim.12656>.
- [36] J. Kocoń et al., ”ChatGPT: Jack of all trades, master of none”, *Information Fusion*, vol. 99, s. 101–861, 2023. DOI: <https://doi.org/10.1016/j.inffus.2023.101861>.
- [37] A. Radford, K. Narasimhan, T. Salimans ja I. Sutskever, ”Improving Language Understanding by Generative Pre-Training”, 2018. url: [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf).
- [38] J. Devlin, M.-W. Chang, K. Lee ja K. Toutanova, ”BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, teoksessa *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran ja T. Solorio, toim., Minneapolis, Minnesota: Association for Computational Linguistics, kesäkuu 2019, s. 4171–4186. DOI: <https://doi.org/10.18653/v1/N19-1423>. url: <https://aclanthology.org/N19-1423/>.
- [39] C. Sun, A. Myers, C. Vondrick, K. Murphy ja C. Schmid, ”VideoBERT: A Joint Model for Video and Language Representation Learning”, teoksessa *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, s. 7463–7472. DOI: <https://doi.org/10.1109/ICCV.2019.00756>.
- [40] H. Ali, E. Hashmi, S. Yayilgan Yildirim ja S. Shaikh, ”Analyzing Amazon Products Sentiment: A Comparative Study of Machine and Deep Learning, and Transformer-Based Techniques”, *Electronics*, vol. 13, nro 7, 2024. DOI: <https://doi.org/10.3390/electronics13071305>.

- [41] E. Hashmi ja S. Y. Yayilgan, "A robust hybrid approach with product context-aware learning and explainable AI for sentiment analysis in Amazon user reviews", *Electronic Commerce Research*, vol. 25, nro 6, s. 5139–5171, 2025. DOI: <https://doi.org/10.1007/s10660-024-09896-5>.
- [42] P. Rasappan, M. Premkumar, G. Sinha ja K. Chandrasekaran, "Transforming sentiment analysis for e-commerce product reviews: Hybrid deep learning model with an innovative term weighting and feature selection", *Information Processing Management*, vol. 61, nro 3, s. 103654, 2024. DOI: <https://doi.org/10.1016/j.ipm.2024.103654>.
- [43] N. Gozuacik, C. O. Sakar ja S. Ozcan, "Social media-based opinion retrieval for product analysis using multi-task deep neural networks", *Expert Systems with Applications*, vol. 183, s. 115388, 2021. DOI: <https://doi.org/10.1016/j.eswa.2021.115388>.
- [44] Z. Wang, P. Gao ja X. Chu, "Sentiment analysis from Customer-generated online videos on product review using topic modeling and Multi-attention BLSTM", *Advanced Engineering Informatics*, vol. 52, s. 101588, 2022. DOI: <https://doi.org/10.1016/j.aei.2022.101588>.
- [45] J. Joung ja H. M. Kim, "Explainable neural network-based approach to Kano categorisation of product features from online reviews", *International Journal of Production Research*, vol. 60, nro 23, s. 7053–7073, 2022. DOI: <https://doi.org/10.1080/00207543.2021.2000656>.