Rīgas 64. Vidusskola

**Noskaņojuma Analīze YouTube Komentāriem**

Zinātniski pētnieciskais darbs: Datorzinātņu un informātikas sadaļā

**Darba autors:**

Rīgas 64. vidusskolas 12. klases skolnieks

Kristers Laganovskis

**Darba vadītājs:**

Rīgas 64. vidusskolas programmēšanas skolotājs

Edvards Bukovskis

Rīga 2023

**Anotācija**

Zinātniski pētniecisko darbu: “Noskaņojuma Analīze YouTube Komentāriem” Izstrādājis autors Rīgas 64. vidusskolas 12.DIT klases skolēns – Kristers Laganovskis. Darba Vadītājs Rīgas 64. vidusskolas programmēšanas skolotājs - Edvards Bukovskis.

Zinātniski pētnieciskais darbs tika veikts ar mērķi, izpētīt kā mašīnmācīšanos, dabiskās valodas apstrādi (NLP), iegūt noskaņojuma analīzi. Kā, tiek veikta efektīva datu ievākšana. Kā arī, kā tiek veikta efektīva datu apstrāde. Zinātniski pētnieciskajā darbā balstoties uz analizēto un apkopoto informāciju par dabiskās valodas apstrādi, datu ievākšanu, kā arī to apstrādi, tika izveidots mājaslapas paplašinājums, ar kura palīdzību spēja analizēt YouTube komentāru noskaņojumu attiecībā uz video saturu.

Teorētiskajā daļa tika detalizēti izpētīti dabiskās valodas apstrādes (NLP) pamati, teksta noskaņojuma analizēšanai. Tās pielietošana sociālajā un ētiskajā kontekstā, kā arī tās nozīme mūsdienu un vēsturiskajā kontekstā. Kā, tiek veikta efektīva datu ievākšana un tā principi. Kā arī datu apstrādes principi, lai nodrošinātu maksimālu precizitāti noskaņojuma novērtējumu.

Praktiskā daļā tika izstrādāts mājaslapas paplašinājums, kas sniedz iespēju vizualizēt statistiskus datus par YouTube komentāru noskaņojumu, sniedzot vērtīgu perspektīvu par to, kāda ir publikas reakcija uz konkrēto video saturu. Paplašinājums nodrošina iespēju efektīvi filtrēt komentārus pēc to noskaņojuma, padarot analīzes procesu, vēl precīzāku un efektīvāku.

Atslēgas Vārdi: Datu ievākšana, datu apstrāde, Komentāri, noskaņojuma analīze, mājaslapas paplašinājums, dabiskās valodas apstrāde (NLP), mašīnmācīšanās.

**Abstract**

The scientific research work entitled "Analysis of Sentiment in YouTube Comments" was conducted by Kristers Laganovskis, a student of Class 12.DIT at Riga Secondary School No. 64. The research was supervised by Edvards Bukovskis, a programming instructor at Riga Secondary School No. 64.

The aim of the scientific research was to explore machine learning, natural language processing (NLP), and sentiment analysis. It delved into effective data collection methods and efficient data processing techniques. Drawing upon the analyzed and compiled information regarding natural language processing, data acquisition, and processing, an extension for a website was developed. This extension facilitated the analysis of sentiment in YouTube comments in relation to video content.

The theoretical section meticulously examined the fundamentals of natural language processing (NLP) and sentiment analysis of text. It discussed their application in social and ethical contexts, as well as their significance in contemporary and historical contexts. The research also addressed the principles of effective data gathering and its methodologies, along with data processing principles aimed at ensuring maximum precision in sentiment analysis.

In the practical section, an extension for the website was developed to provide the ability to visualize statistical data regarding sentiment in YouTube comments. This offered valuable insights into the audience's reaction to specific video content. The extension enabled efficient filtering of comments based on their sentiment, thereby refining the analysis process for greater precision and effectiveness.

Keywords: Data collection, data processing, comments, sentiment analysis, website extension, natural language processing (NLP), machine learning.

**Saturs**

Anotācija….....................................................................................................................................1

Annotation.....................................................................................................................................1

Ievads…………………………………………………………………………………………………………………………………..…..1

Secinājumi………………………………………………………………………………………………………………….……………..1

Izmantotie informācijas avoti…………………………………………………………………………………………………….1

Pielikumi……………………………………………………………………………………………………………………………………1

**Ievads**

Mūsdienās informācijas klāsts ir liels un apjomīgs, taču tas var radīt savas neērtības gan kā satura patērgātājiem, gan kā satura veidotājiem. Problēma ir meklējama tajā, ka cilvēki bieži vien pārāk vērš lielu uzticību satura veidotājam, kas ne vienmēr var būt labi, jo satura veidotājs var speciāli apslēpt un manipulēt ar to patērētājiem. Tāpēc ir ļoti svarīgi zināt, ko citi cilvēki domā par noskatīto video saturu. Salīdzinoši nesen 2021. gada. Novembrī YouTube noņēma iespēju redzēt “nepatīk” balsu skaitu, kas ietekmēja patērētājus negatīvi. Šāds lēmums veicināja nekorektu novērtējumu par video saturu, jo neuzrādoties “nepatīk” balsu skaitam, ir grūti spriest, cik cilvēki ir negatīvi novērtējuši video saturu. Tāpēc viens no veidiem, kā noskaidrot, ko cilvēki domā par video saturu ir vērsties pie komentāru sadaļas, taču tad atkal rodas problēma. Ja tie ir desmit komentāri, tas nerada lielas neērtības, bet, ja tie ir simti vai tūkstoši, tas noteikti rada neērtības, jo šādu apjomu izskatīt aizņems ļoti ilgu laiku.

Šajā darbā tiks pētīta YouTube video komentāru sadaļa, kā ar šiem datiem spēj nolasīt noskaņojumu par video saturu ar dabiskās valodas apstrādes (NLP) modeli, kādas metodes tiek pielietotas, lai iegūtu nepieciešamos datus analīzei, kā arī kādā veidā dati ir jāapstrādā, lai veiksmīgi iegūtu noskaņojuma analīzi. Tiks veidots arī paplašinājums ar kura palīdzību ikviens varēs analizēt komentārus, kā arī tos filtrēt pēc noteiktām kategorijām, kas ir noderīgi ne tikai satura uztvērējam, bet arī satura veidotājam attiecīgi ietaupot laiku filtrējot komentārus.

Viens no mašīnmācīšanās pielietojumiem ir dabiskās valodas apstrādes (NLP), to izmanto, lai datori varētu saprast un interpretēt cilvēku rakstīto vai runāto valodu. Tas ietver teksta analīzi, tulkošanu, noskaņojuma analīzi, teksta ģenerēšanu un daudzas citas funkcijas, kas atvieglo komunikāciju starp cilvēkiem un datoriem. Šis pētījums koncentrēsies uz noskaņojuma analīzi. Dabiskās valodas apstrāde atver durvis daudzām inovatīvām lietojumprogrammām, tai skaitā virtuālajiem asistentiem, automātiskai tulkotājiem un daudz ko citu. Šis tehnoloģijas virziens strauji attīstās, sniedzot plašas iespējas dažādām nozarēm, tostarp medicīnai, izglītībai un komunikācijas nozarēm.

**Darba mērķis:** Iepazīties un izpētīt dabiskās valodas apstrādes modeli – NLP, noskaņojuma nolasīšanai. Izpētīt kādā veidā ir jāapstrādā teksts, lai no tā iegūtu precīzāko novērtējumu. Ievākt efektīvi YouTube komentārus, lai pēc tam tos apstrādātu. Izveidot paplašinājumu, kas ļaus analizēt un filtrēt komentārus, sniedzot vērtīgu informāciju par redzētu saturu video.

**Hipotēze:** Izmantojot NLP modeli, ir iespējams veiksmīgi analizēt un interpretēt YouTube video komentārus, lai novērtētu skatītāju noskaņojumu par video saturu.

**Darba uzdevumi:**

1. Izpētīt NLP modeli un kā to pielietot noskaņojuma nolasīšanai
2. Izpētīt, kā iegūt nepieciešamos datus no YouTube video komentāru sadaļas
3. Izpētīt, kā apstrādāt datus efektīvi, lai tie rezultētos precīzā novērtējumā
4. Izstrādāt paplašinājumu, kas ļauj ikvienam analizēt un filtrēt komentārus pēc noteiktām kategorijām.
5. Veikt secinājumus

**Izmantotās darba metodes:** Literatūras analīze, lai izpētītu NLP modeli noskaņojuma noteikšanai. NLP (Dabiskās valodas apstrādes) modeļu pielietojums komentāru analīzei un noskaņojuma nolasīšanai. Datu iegūšana no YouTube video komentāru sadaļas ar API (lietojumprogrammas saskarnes) palīdzību. Statistikas un analīzes metodes, lai interpretētu rezultātus un veiktu secinājumus. Atvērtā pirmkoda pielietošana, “Python” programmēšanas valodas pielietošana radošās daļas veikšanai.

**Darba struktūra:** Darbs sastāv no x nodaļām, x apakšnodaļām, secinājumiem, izmantoto literatūras avotu saraksta un x pielikumiem. Darbā ir x attēli.