# РЕФЕРАТ

**Актуальність теми:**

1. покращення якості визначення емоцій в повідомленнях написаних українською мовою;
2. ефективний інструмент моніторингу і оцінювання тематики груп/каналів

**Мета дослідження** дисертаційної роботи є покращення сентимент аналізу у текстових потоках даних.

Розроблені методи і програмні засоби повинні відповідати таким вимогам:

* більш висока в порівнянні з існуючими моделями якість запропонованих методів;
* адаптація методів для підтримки текстів різних мов. В даній дисертаційній роботі розглядаються тексти користувачів українською та англійською мовами;
* візуалізація результатів аналізу.

Для реалізації поставленої мети були сформульовані **наступні завдання**:

1. провести аналіз існуючих методів та програмних рішень;
2. обґрунтувати вибір методу сентимент аналізу;
3. розробити математичну модель класифікації повідомлень користувачів за сентиментом;
4. покращити методи аналізу емоційної складової для потоків даних;
5. реалізувати запропоновані методи у вигляді програмного засобу. візуалізувати аналіз емоційної складової потоку за певний період даних та провести експериментальні дослідження з метою визначення ефективності роботи методів і моделей.

**Об'єкт і предмет дослідження.** Об'єктом дослідження є потік повідомлень користувачів, представлені у вигляді потоку неструктурованих текстів на природній мові.

**Методи дослідження:** У даній дисертаційній роботі застосовувалися методи обробки природної мови, засновані на правилах, словниках та існуючих лінгвістичних ресурсах, і ймовірнісних тематичних моделях, заснованих на комплексі методів машинного навчання.

**Наукова новизна:** У даній роботі запропоновані поліпшення для методів визначення сентименту в текстах, написаних українською мовою, представлених у вигляді потоку неструктурованих текстових даних природньою мовою, з використання алгоритму машинного навчання з учителем.

**Практичне значення отриманих результатів** визначається тим, що запропоноване покращення до алгоритму може бути використано для прикладного аналізу емоційної складової контенту текстових каналів, груп в меcенджері telegram і т.д..

**Зв’язок роботи з науковими програмами, планами, темами:** робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Методи та технології високопродуктивних обчислень та обробки надвеликих масивів даних». Державний реєстраційний номер 0117U000924.

**Публікації:** Наукові положення дисертації опубліковані в тезах конференції *«ІНФОРМАТИКА ТА ОБЧИСЛЮВАЛЬНА ТЕХНІКА – ІОТ-2019»*

**Ключові слова**: СЕМАНТИЧНИЙ АНАЛІЗ, ПОТІК ДАНИХ, GRADIENT BOOSTING TREE, АНАЛІЗ ТОНАЛЬНОСТІ