

ΔΡΑΣΗ ΕΘΝΙΚΗΣ ΕΜΒΕΛΕΙΑΣ

ΕΡΕΥΝΩ - ΔΗΜΙΟΥΡΓΩ - ΚΑΙΝΟΤΟΜΩ

«ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ, ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ & ΚΑΙΝΟΤΟΜΙΑ» (ΕΠΑνΕΚ)
Ειδική Υπηρεσία Διαχείρισης Επιχειρησιακού Προγράμματος Ανταγωνιστικότητα
Επιχειρηματικότητα και Καινοτομία (ΕΥΔ ΕΠΑνΕΚ)
Ειδική Υπηρεσία Διαχείρισης και Εφαρμογής Δράσεων στους τομείς
Έρευνας, Τεχνολογικής Ανάπτυξης και Καινοτομίας
(ΕΥΔΕ ΕΤΑΚ)

**ΕΡΓΟ: Ολοκληρωμένη Πλατφόρμα Blockchain και Έξυπνη Έφαρμολή
Κινητής Συσκευής στη Μάχη Ενάντια στη Σπατάλη Τροφίμων προς
Όφελος των Επισιτιστικά Ανασφαλών Πολιτών**

ΚΩΔΙΚΟΣ ΕΡΓΟΥ: Τ2ΕΔΚ- 05051

ΠΑΡΑΔΟΤΕΑ

**Π9: Πρώτη έκδοση του υποσυστήματος συλλογής και ανάλυσης δεδομένων
μεγάλου όγκου**

Π10: Εργαλεία ανάλυσης κοινωνικών δικτύων.

**Π11: Τελική έκδοση του υποσυστήματος συλλογής και ανάλυσης δεδομένων
μεγάλου όγκου**

Διάρκεια Έργου	30/11/2021 ως 29/11/2023
Ημερομηνία Παράδοσης	30/09/2023
Υπεύθυνος Παραδοτέου:	ΕΜΠ
Επίπεδο Διάχυσης:	Ιδιωτικό
Τρέχουσα Κατάσταση Παραδοτέου:	Τελικό προς έλεγχο
Έκδοση:	Π9-Π10-Π11 F
Ιστοσελίδα Έργου:	https://blockfoodwaste.eu/

ΦΟΡΕΙΣ ΥΛΟΠΟΙΗΣΗΣ:



Cyclefi



isMOOD

ΑΘΗΝΑ

ΣΕΠΤΕΜΒΡΙΟΣ 2023

Περιεχόμενα

1. Σύνοψη προς τη Διαχειριστική Αρχή (Executive Summary).....	6
2. Εισαγωγή.....	10
2.1. Στόχοι του Συστήματος	11
2.2. Δήλωση Προβλήματος	12
3. Τεχνολογίες Συστήματος (Technology Stack).....	15
3.1. Python - Scripting και Επεξεργασία Δεδομένων.....	15
3.2. Twitter API και Reddit API - Συλλογή δεδομένων.....	16
3.3. MongoDB - Αποθήκευση Δεδομένων	17
3.4. Dash Plotly - Δημιουργία Dashboard	19
3.5. Άλλες Βιβλιοθήκες και Εργαλεία	20
3.5.1. NLTK (Εργαλειοθήκη φυσικής γλώσσας)	20
3.5.2. GenSim	21
3.5.3. PyLDAvis	21
4. Αρχιτεκτονική Συστήματος	22
4.1. Docker Deployment	26
5. Σχεδιασμός Μονάδων Συστήματος.....	28
5.1. Μονάδα Συλλογής Δεδομένων (Twitter & Reddit)	28
5.2. Μονάδα Προεπεξεργασίας Δεδομένων	29
5.2.1. Μετάφραση κειμένου στην Αγγλική Γλώσσα	29
5.2.2. Μετατροπή κειμένου σε πεζά.....	29
5.2.3. Αφαίρεση Ετικετών HTML.....	29
5.2.4. Επέκταση Συντομογραφιών	30
5.2.5. Μετατροπή αριθμών σε λέξεις	30
5.2.6. Αφαίρεση ειδικών χαρακτήρων.....	30
5.2.7. Αφαίρεση λευκών διαστημάτων.....	30
5.2.8. Tokenization Word	31
5.2.9. Θεματοποίηση και Λημματοποίηση	31

5.2.10.	Αφαίρεση κοινών λέξεων.....	31
5.3.	Μονάδα συστήματος βάσης δεδομένων - Ρύθμιση και σχήμα MongoDB	32
5.3.1.	Εγκατάσταση και διαμόρφωση.....	32
5.3.2.	Σχεδιασμός Σχήματος.....	32
5.4.	Μονάδα οπτικοποίησης των αποτελεσμάτων με τη χρήση του Dash Plotly	33
5.4.1.	Βασικές Οπτικοποιήσεις	34
5.4.2.	Βασικές Λειτουργίες.....	34
5.5.	Σχεδιασμός της στρατηγικής προσωρινής αποθήκευσης (cache DB)	35
5.5.1.	Τεχνική Υλοποίηση	36
6.	Αναλύσεις.....	37
6.1.	Ανάλυση Συναισθήματος με χρήση του VADER	37
6.1.1.	Προεπεξεργασία	37
6.1.2.	Βαθμολογία Συναισθημάτων	37
6.1.3.	Ερμηνεία.....	37
6.1.4.	Πλεονεκτήματα του VADER στην ανάλυση των μέσων κοινωνικής δικτύωσης. 38	
6.1.5.	Εφαρμογή στο Έργο	38
6.2.	Ανάλυση Συχνότητας.....	38
6.2.1.	Μεθοδολογία Ανάλυσης Συχνότητας	38
6.2.2.	Ανάλυση και Ερμηνεία	39
6.2.3.	Εφαρμογή στο Έργο	39
6.3.	Μοντελοποίηση θεμάτων και λανθάνουσα κατανομή Dirichlet (LDA)	39
6.3.1.	Περιγραφή της Μεθόδου.....	39
6.3.2.	Σημασία στο Πλαίσιο των Μεγάλων Δεδομένων	40
6.3.3.	Εφαρμογή στο Έργο	40
7.	Επισκόπηση Δυναμικού Πίνακα Ελέγχου	41
7.1.	Συνιστώσες.....	41
7.2.	Επεξεργασία Δεδομένων	41
7.3.	Ανάλυση Συναισθήματος	42
7.4.	Συναίσθημα με την πάροδο του χρόνου	43
7.5.	Ανάλυση Hashtags και Λέξεων.....	44

7.6.	Δίκτυο συνεμφάνισης Hashtags.....	45
7.7.	Συχνότητα λέξεων έναντι συναισθήματος.....	47
7.8.	Μοντελοποίηση Θεμάτων	48
8.	Συμπεράσματα	52
9.	Βιβλιογραφικές Αναφορές	55

Περιεχόμενα Σχημάτων

Σχήμα 1-1: Github Repository - BLOCKFOODWASTE Υποσύστημα Συλλογής & Ανάπτυξης Δεδομένων	6
Σχήμα 1-2: UI Συστήματος - BLOCKFOODWASTE Υποσύστημα Συλλογής & Ανάπτυξης Δεδομένων	8
Σχήμα 4-1: Αρχιτεκτονική Συστήματος (Conceptual Diagram)	22
Σχήμα 4-2: Διάγραμμα UML - Sequence	24
Σχήμα 4-3: Διάγραμμα Αλληλοεπιδράσεων (interactions)	25
Σχήμα 4-4: Διάγραμμα Components Interaction	25
Σχήμα 7-1: Κατανομή Συναισθήματος	42
Σχήμα 7-2: Συναίσθημα με την πάροδο του χρόνου	43
Σχήμα 7-3: Κατάταξη των hashtags	44
Σχήμα 7-4: Δίκτυο Συνεμφάνισης των hastags	46
Σχήμα 7-5: Συχνή χρήση λέξεων έναντι συναισθήματος	47
Σχήμα 7-6: Μοντελοποίηση Θεμάτων	49

1. Σύνοψη προς τη Διαχειριστική Αρχή (Executive Summary)

Το παρόν ‘φυσικό’ παραδοτέο περιλαμβάνει το σύνολο των εργασιών που έλαβαν χώρα για την ανάπτυξη του υποσυστήματος συλλογής και ανάλυσης δεδομένων από τα κοινωνικά δίκτυα με τη χρήση αλγορίθμων NLP (Natural Language Processing) για την ανίχνευση και ερμηνεία προθέσεων και διαθέσεων των χρηστών (sentiment mining) σε σχέση με το φαινόμενο της σπατάλης τροφίμων.

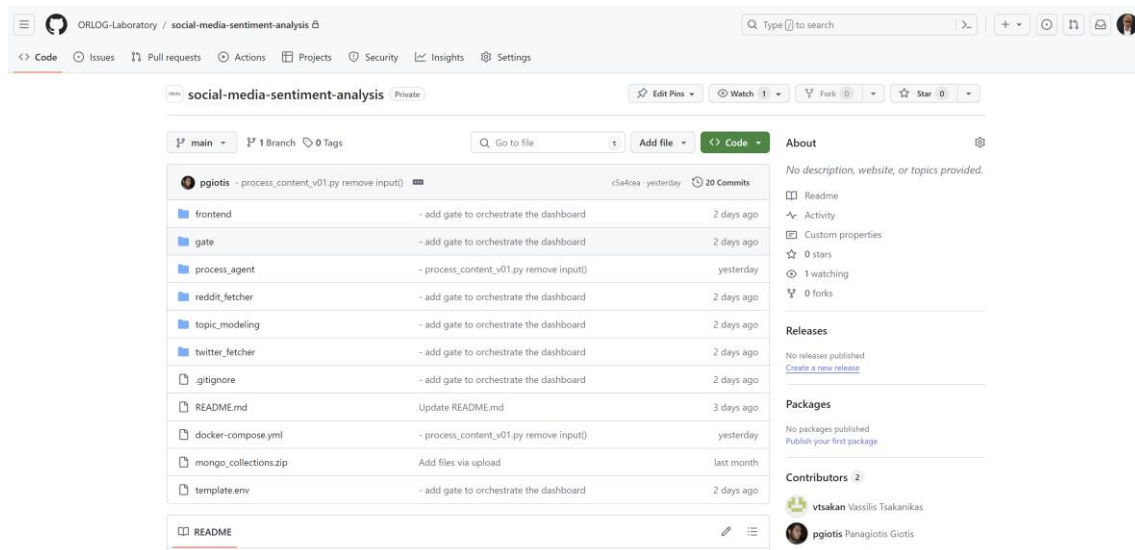
Στην ουσία το ‘φυσικό’ αυτό παραδοτέο περιλαμβάνει τις εργασίες τριών παραδοτέων και πιο συγκεκριμένα των:

Π9: Πρώτη έκδοση του υποσυστήματος συλλογής και ανάλυσης δεδομένων μεγάλου όγκου

Π10: Εργαλεία ανάλυσης κοινωνικών δικτύων.

Π11: Τελική έκδοση του υποσυστήματος συλλογής και ανάλυσης δεδομένων μεγάλου όγκου

Και τα τρία αυτά παραδοτέα, αναφέρονται στο τεχνικό παράρτημα ως ‘Λογισμικό’. Κατά συνέπεια η τυπική παράδοση τους γίνεται μέσα από τις εργασίες ανάπτυξης που κανείς μπορεί να ελέγξει στην ακόλουθη online πλατφόρμα συνεργατικής ανάπτυξης και φιλοξενίας κώδικα που βρίσκεται στην ακόλουθη διεύθυνση: <https://github.com/ORLOG-Laboratory/social-media-sentiment-analysis> και φαίνεται στο Σχήμα 1.1 που ακολουθεί:



Σχήμα 1-1: Github Repository - BLOCKFOODWASTE Υποσύστημα Συλλογής & Ανάπτυξης Δεδομένων

Επιπρόσθετα, κανείς μπορεί να ελέγξει τη χρήση του UI του εργαλείου που φιλοξενείται στην πλατφόρμα Digital Ocean στην ακόλουθη διεύθυνση: <https://foodwaste.orlog.gsoft.gr/>, όπως φαίνεται στο Σχήμα που ακολουθεί.

Social Media Analysis Dashboard

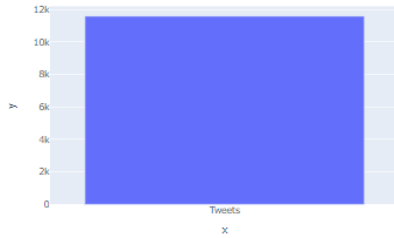
Refresh Data

2023-01-01 → 2024-01-24

Select Post Source:

Both

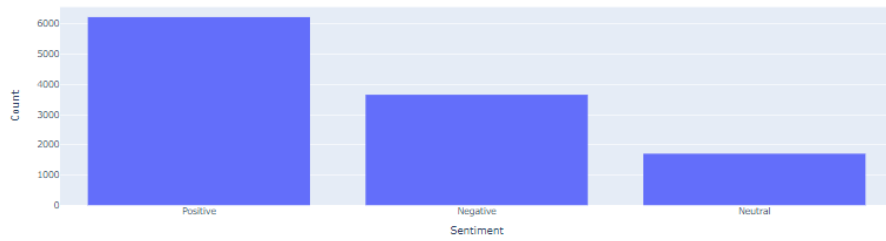
Total Number of Posts



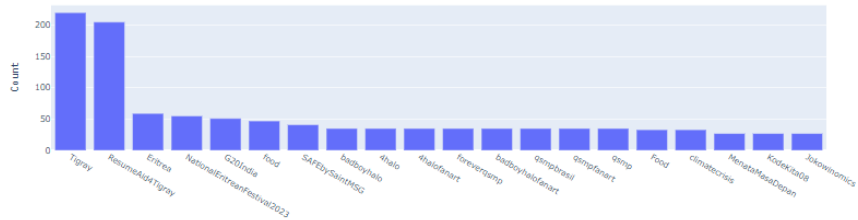
Total Number of Unique Words



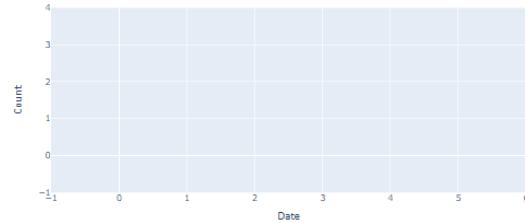
Sentiment Distribution



Top 20 Hashtags



Hashtag Popularity Over Time



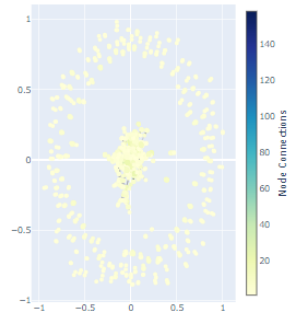
Hashtag



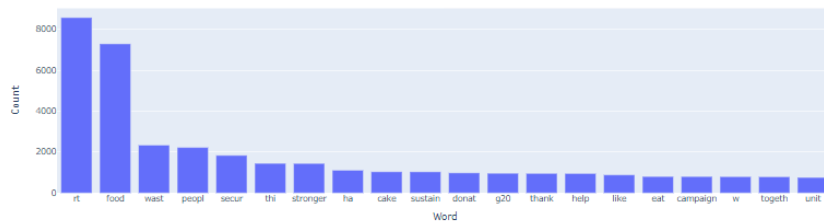


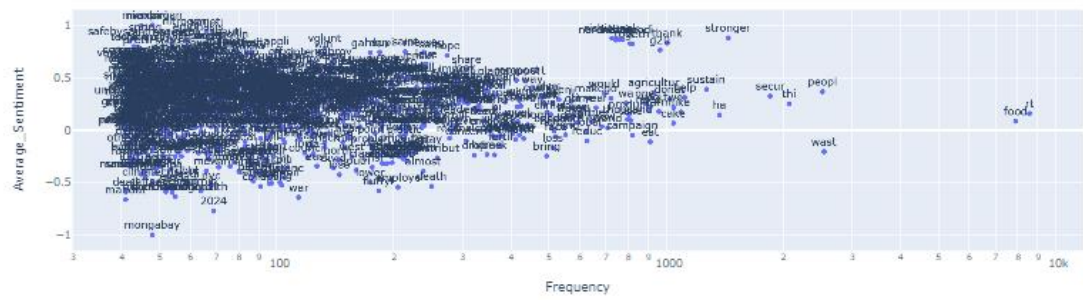




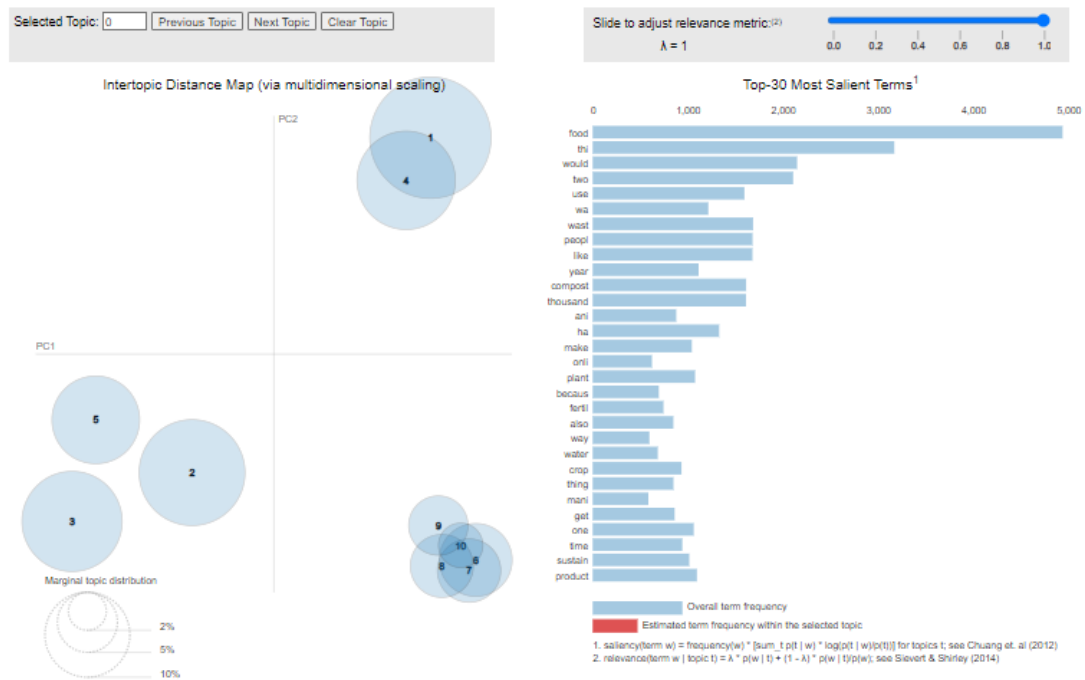


Top 20 Words Across All Tweets





Topic Modeling



Σχήμα 1-2: UI Συστήματος - BLOCKFOODWASTE Υποσύστημα Συλλογής & Ανάπτυξης
Δεδομένων

Παρόλα αυτά, η ομάδα έργου κρίνει σκόπιμο και ουσιαστικό να παραδώσει ταυτόχρονα και το παρόν τεύχος που περιλαμβάνει τις εργασίες των **παραδοτέων Π9, Π10 και Π11** και περιγράφει την αρχιτεκτονική του συστήματος, τη διαδρομή της ανάκτησης δεδομένων από τα επιλεγθέντα κοινωνικά δίκτυα (Twitter & Reddit) και τις οδηγίες χρήσης του αναπτυχθέντος λογισμικού.

2. Εισαγωγή

Στις ενότητες που ακολουθούν, παρουσιάζονται οι εργασίες που έλαβαν χώρα στο πλαίσιο υλοποίησης των **Παραδοτέων Π9, Π10 και Π11**, και οδήγησαν στην ανάπτυξη ενός ολοκληρωμένου συστήματος για τη συλλογή, επεξεργασία, ανάλυση και οπτικοποίηση δεδομένων μέσω κοινωνικής δικτύωσης, εστιάζοντας συγκεκριμένα σε περιεχόμενο που σχετίζεται με τη σπατάλη τροφίμων. Χρησιμοποιώντας τα API μεγάλων πλατφορμών μέσω κοινωνικής δικτύωσης όπως το Twitter και το Reddit, το σύστημα συγκεντρώνει σχετικές αναρτήσεις και συζητήσεις, παρέχοντας πληροφορίες για την αντίληψη και την ευαισθητοποίηση του κοινού σχετικά με τη σπατάλη τροφίμων.

Σε μια εποχή όπου τα μέσα κοινωνικής δικτύωσης διαδραματίζουν κεντρικό ρόλο στη διαμόρφωση της κοινής γνώμης και του λόγου, το αναπτυχθέν σύστημα αξιοποιεί αυτές τις πλατφόρμες αντλώντας και αναλύοντας δεδομένα για ένα από τα σημαντικότερα σύγχρονα προβλήματα. Ο πρωταρχικός στόχος είναι να αναλυθεί ο τρόπος με τον οποίο τα απόβλητα τροφίμων αντιλαμβάνονται και συζητούνται από τους χρήστες, κάτι που μπορεί να προσφέρει πολύτιμες γνώσεις σε οργανισμούς, υπεύθυνους χάραξης πολιτικής και περιβαλλοντολόγους που εργάζονται για τη μείωση της σπατάλης τροφίμων και την προώθηση βιώσιμων πρακτικών.

Στον πυρήνα του συστήματος βρίσκεται ένας ισχυρός αγωγός επεξεργασίας (pipeline) δεδομένων που υλοποιείται στην Python. Αυτός ο αγωγός περιλαμβάνει βήματα για τη συλλογή δεδομένων, την προεπεξεργασία (όπως κανονικοποίηση, αφαίρεση κοινών λέξεων και θεματοποίηση κειμένου), την απόδοση δεικτών και οργάνωση των δεδομένων με βάση τις σύγχρονες αρχές οργάνωσης κειμένων (π.χ. μοντέλο Διανυσματικού Χώρου, ανάλυση συχνότητας), την ανάλυση συναισθήματος και προηγμένες τεχνικές ανάλυσης δεδομένων, όπως ανάλυση συχνότητας και μοντελοποίηση/εξαγωγή θεμάτων (topics modeling/extraction) χρησιμοποιώντας εργαλεία επεξεργασίας φυσικής γλώσσας (NLP), όπως το NLTK και το Gensim. Τα επεξεργασμένα δεδομένα αποθηκεύονται στο MongoDB, μια βάση δεδομένων NoSQL, που επιλέχθηκε για την επεκτασιμότητα και την ευελιξία της στο χειρισμό μεγάλων όγκων μη δομημένων δεδομένων.

Ένα από τα βασικά χαρακτηριστικά αυτού του συστήματος είναι η εφαρμογή ενός μηχανισμού προσωρινής αποθήκευσης για τη βελτίωση της απόδοσης του πίνακα ελέγχου οπτικοποίησης δεδομένων. Κατασκευασμένος χρησιμοποιώντας το Dash Plotly, ο πίνακας εργαλείων παρέχει μια διαδραστική διεπαφή για τους χρήστες να εξερευνούν και να κατανοούν τις τάσεις, τα συναισθήματα και τα θέματα που επικρατούν στη συζήτηση των μέσων κοινωνικής δικτύωσης για τη σπατάλη τροφίμων. Η στρατηγική προσωρινής αποθήκευσης μειώνει σημαντικά τον χρόνο φόρτωσης για πολύπλοκες απεικονίσεις, βελτιώνοντας έτσι την εμπειρία του χρήστη.

Ολόκληρο το σύστημα εφαρμόζεται σε container χρησιμοποιώντας το Docker, διασφαλίζοντας εύκολη ανάπτυξη, επεκτασιμότητα και συνέπεια σε διαφορετικά περιβάλλοντα. Αυτή η προσέγγιση όχι μόνο εξορθολογίζει τη διαδικασία ανάπτυξης αλλά ενισχύει επίσης την αξιοπιστία και την επεκτασιμότητα του συστήματος.

Το σύστημα αντιμετωπίζει ένα σημαντικό περιβαλλοντικό ζήτημα αξιοποιώντας μεγάλα δεδομένα από τα μέσα κοινωνικής δικτύωσης, προσφέροντας πολύτιμες γνώσεις για την κοινή γνώμη και τη συμπεριφορά σχετικά με τα απόβλητα τροφίμων. Τα ευρήματα από αυτήν την ανάλυση μπορούν να ενημερώσουν τους υπεύθυνους χάραξης πολιτικής, τις περιβαλλοντικές οργανώσεις και το ευρύ κοινό, συμβάλλοντας σε πιο αποτελεσματικές στρατηγικές για τη μείωση της σπατάλης τροφίμων και την προώθηση της βιωσιμότητας. Το αναπτυχθέν σύστημα καταδεικνύει τη δύναμη του συνδυασμού των αναλύσεων των μέσων κοινωνικής δικτύωσης με τις σύγχρονες τεχνολογίες επεξεργασίας δεδομένων και οπτικοποίησης για τη δημιουργία πρακτικών πληροφοριών για κρίσιμα παγκόσμια ζητήματα.

2.1. Στόχοι του Συστήματος

Οι πρωταρχικοί τεχνολογικοί στόχοι του συστήματος συλλογής και ανάλυσης δεδομένων μεγάλου όγκου αναλύονται σε έξι (6) βασικές διαστάσεις, πιο συγκεκριμένα τις ακόλουθες:

1. Συλλογή και συγκέντρωση δεδομένων: Συστηματική συλλογή και συγκέντρωση σχετικών δεδομένων από το Twitter και το Reddit, εστιάζοντας σε αναρτήσεις, σχόλια και συζητήσεις που σχετίζονται με τη σπατάλη τροφίμων.

2. Προεπεξεργασία και διαχείριση δεδομένων: Εφαρμογή ισχυρών τεχνικών προεπεξεργασίας για τον καθαρισμό και τη δομή των συλλεγόμενων δεδομένων, καθιστώντας τα κατάλληλα για ανάλυση. Αυτό περιλαμβάνει την κανονικοποίηση κειμένου, την θεματοποίηση, τη δεικτοδότηση κειμένου και άλλες τεχνικές Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing).

3. Ανάλυση συναισθήματος: Ανάλυση συναισθήματος των αναρτήσεων στα μέσα κοινωνικής δικτύωσης προκειμένου να σφυγμομετρηθεί η κοινή γνώμη σχετικά με τη σπατάλη τροφίμων, προσδιορίζοντας εάν τα συναισθήματα είναι θετικά, αρνητικά ή ουδέτερα.

4. Μοντελοποίηση θεμάτων και ανάλυση τάσεων: Εντοπισμός βασικών θεμάτων (Topic Extraction) στο περιεχόμενο των μέσων κοινωνικής δικτύωσης που σχετίζονται με τη σπατάλη τροφίμων και ανάλυση των τάσεων με την πάροδο του χρόνου.

5. Διαδραστική Οπτικοποίηση Δεδομένων: Ανάπτυξη ενός διαδραστικού πίνακα εργαλείων χρησιμοποιώντας το Dash Plotly, που επιτρέπει στους χρήστες να εξερευνούν τα δεδομένα και τις πληροφορίες με διαισθητικό και προσιτό τρόπο.

6. Δημιουργία Insight and Reporting: Δημιουργία πρακτικών πληροφοριών που μπορούν να ενημερώσουν τους ενδιαφερόμενους φορείς, συμπεριλαμβανομένων περιβαλλοντικών ομάδων, υπευθύνων χάραξης πολιτικής και το ευρύ κοινό, σχετικά με την τρέχουσα κατάσταση της ευαισθητοποίησης και της στάσης του κοινού απέναντι στα απόβλητα τροφίμων.

Η επίτευξη των παραπάνω τεχνολογικών στόχων, μεταφράζεται σε μια σειρά από στόχους υψηλού επιπέδου, με ιδιαίτερη σημασία για το έργο BLOCKFOODWASTE και τον αντίκτυπο του σε μια σειρά σημαντικά θέματα, όπως:

- **Περιβαλλοντική ευαισθητοποίηση:** Αναλύοντας τη συζήτηση των μέσων κοινωνικής δικτύωσης, το αναπτυχθέν σύστημα υπογραμμίζει την ευαισθητοποίηση του κοινού και τις ανησυχίες σχετικά με τη σπατάλη τροφίμων, συμβάλλοντας στις προσπάθειες διατήρησης του περιβάλλοντος.
- **Πολιτική και λήψη αποφάσεων:** Οι γνώσεις που προκύπτουν από την ανάλυση μπορούν να ενημερώσουν τους υπεύθυνους χάραξης πολιτικής και τους οργανισμούς στη δημιουργία στοχευμένων εκστρατειών και πολιτικών για τη μείωση της σπατάλης τροφίμων και την προώθηση βιώσιμων πρακτικών.
- **Ακαδημαϊκή Συνεισφορά:** Το αναπτυχθέν σύστημα καταδεικνύει την ανάγκη εφαρμογής μιας σειράς σύγχρονων υπολογιστικών μεθόδων και μοντέλων (π.χ. συγκέντρωση, διαχείριση και ανάλυση μεγάλων δεδομένων, νέοι μέθοδοι επεξεργασίας φυσικής γλώσσας) σε περιβαλλοντικές μελέτες, συμβάλλοντας στην ακαδημαϊκή έρευνα σε αυτόν τον τομέα.
- **Δημόσια Δέσμευση:** Ο διαδραστικός πίνακας εργαλείων χρησιμεύει ως εργαλείο για τη δημόσια συμμετοχή, εκπαιδύοντας τους χρήστες σχετικά με θέματα σχετικά με τα απορρίμματα τροφίμων και ενδεχομένως επηρεάζοντας τις συμπεριφορές και τις στάσεις τους.
- **Τεχνολογική Καινοτομία:** Το αναπτυχθέν σύστημα παρουσιάζει την ενσωμάτωση διαφόρων τεχνολογιών και εργαλείων στην ανάλυση μεγάλων δεδομένων και επεξεργασία φυσικής γλώσσας, θέτοντας προηγούμενο για μελλοντική έρευνα σε παρόμοιους τομείς.

Συνολικά, το αναπτυχθέν σύστημα καταδεικνύει τη δύναμη των αναλύσεων των μέσων κοινωνικής δικτύωσης στην κατανόηση και την αντιμετώπιση κρίσιμων περιβαλλοντικών ζητημάτων όπως η σπατάλη τροφίμων, γεφυρώνοντας το χάσμα μεταξύ των γνώσεων που βασίζονται σε δεδομένα και των εφαρμογών του πραγματικού κόσμου.

2.2. Δήλωση Προβλήματος

Το πρόβλημα που προσπαθεί να αντιμετωπίσει το αναπτυχθέν σύστημα συλλογής και ανάλυσης μεγάλων δεδομένων, επικεντρώνεται στην έλλειψη συνολικής κατανόησης της κοινής αντίληψης και των αιτιών εμφάνισης του φαινομένου της σπατάλης τροφίμων, όπως αντικατοπτρίζεται στις πλατφόρμες κοινωνικών μέσων. Παρά την αυξανόμενη ευαισθητοποίηση σχετικά με τη σπατάλη τροφίμων ως κρίσιμο περιβαλλοντικό και κοινωνικό ζήτημα, υπάρχει ένα κενό στην κατανόηση του τρόπου με τον οποίο γίνεται αντιληπτό, συζητείται και αντιμετωπίζεται από το κοινό στις ψηφιακές πλατφόρμες.

Οι βασικές πτυχές του προβλήματος περιλαμβάνουν:

1. Περιορισμένη εικόνα του κοινού αισθήματος: Οι παραδοσιακές μέθοδοι μέτρησης της κοινής γνώμης σχετικά με τα απόβλητα τροφίμων, όπως έρευνες ή συνεντεύξεις, είναι συχνά περιορισμένες σε εύρος και κλίμακα. Τα μέσα κοινωνικής δικτύωσης, με την τεράστια βάση χρηστών τους, προσφέρουν μια πιο εκτεταμένη και ποικιλόμορφη πηγή δεδομένων, αλλά απαιτούν εξελιγμένα εργαλεία για αποτελεσματική ανάλυση.

2. Πολυπλοκότητα των δεδομένων μέσω κοινωνικής δικτύωσης: Οι πλατφόρμες κοινωνικών μέσων παράγουν τεράστιες ποσότητες μη δομημένων δεδομένων, καθιστώντας δύσκολη την εξαγωγή σχετικών και ουσιαστικών πληροφοριών. Τα δεδομένα είναι συχνά θορυβώδη, περιέχουν αργκό, emoji και ακανόνιστη σύνταξη, γεγονός που περιπλέκει την ανάλυση.

3. Δυναμική φύση του διαδικτυακού λόγου: Τα θέματα, οι τάσεις και τα συναισθήματα στα μέσα κοινωνικής δικτύωσης εξελίσσονται συνεχώς. Η αποτύπωση και η ανάλυση αυτών των δυναμικών σε πραγματικό χρόνο ή σε συγκεκριμένες περιόδους είναι ζωτικής σημασίας για την κατανόηση του μεταβαλλόμενου τοπίου της κοινής γνώμης σχετικά με τη σπατάλη τροφίμων.

4. Ανάγκη για Ενημερωτικές Πληροφορίες: Οι ενδιαφερόμενοι, συμπεριλαμβανομένων των περιβαλλοντικών ομάδων, των υπευθύνων χάραξης πολιτικής και των εκπαιδευτικών, απαιτούν συγκεκριμένες γνώσεις για τη διαμόρφωση στρατηγικών και παρεμβάσεων για τη μείωση της σπατάλης τροφίμων. Ωστόσο, η μετατροπή των ακατέργαστων δεδομένων μέσω κοινωνικής δικτύωσης σε πληροφορίες που μπορούν να χρησιμοποιηθούν είναι ένα πολύπλοκο έργο.

Το σύστημα στοχεύει να αντιμετωπίσει αυτές τις προκλήσεις χρησιμοποιώντας προηγμένες τεχνικές συλλογής δεδομένων, προεπεξεργασίας και ανάλυσης. Αναλύοντας το περιεχόμενο των μέσων κοινωνικής δικτύωσης που σχετίζεται με τη σπατάλη τροφίμων, το σύστημα επιδιώκει να αποκαλύψει τις αποχρώσεις της αντίληψης και του συναισθήματος του κοινού, παρέχοντας πολύτιμες γνώσεις που μπορούν να οδηγήσουν στη λήψη τεκμηριωμένων αποφάσεων και αποτελεσματικές επικοινωνιακές στρατηγικές για την καταπολέμηση της σπατάλης τροφίμων.

Πράγματι, τα τελευταία δεκαπέντε -τουλάχιστον- χρόνια, οι πλατφόρμες μέσω κοινωνικής δικτύωσης έχουν αναδειχθεί ως ισχυρά εργαλεία για τη μέτρηση και τον επηρεασμό της κοινής γνώμης σε διάφορα θέματα, συμπεριλαμβανομένης της σπατάλης τροφίμων. Ο ρόλος των μέσων κοινωνικής δικτύωσης στην κατανόηση της αντίληψης του κοινού για τα απόβλητα τροφίμων είναι πολύπλευρος:

- **Ευρεία προσέγγιση χρηστών και διαφορετική βάση χρηστών:** Οι πλατφόρμες μέσω κοινωνικής δικτύωσης διαθέτουν μια τεράστια και ποικιλόμορφη βάση χρηστών, γεγονός που τις καθιστά ιδανικές για την καταγραφή ενός ευρέος φάσματος απόψεων, εμπειριών και στάσεων σχετικά με τη σπατάλη τροφίμων. Αυτή η ποικιλομορφία παρέχει μια πιο ολοκληρωμένη κατανόηση της αντίληψης του κοινού από τις παραδοσιακές μεθόδους έρευνας.
- **Πληροφορίες σε πραγματικό χρόνο:** Τα μέσα κοινωνικής δικτύωσης επιτρέπουν την παρακολούθηση της κοινής γνώμης σε πραγματικό χρόνο. Αυτή η αμεσότητα είναι ζωτικής σημασίας για την κατανόηση της δυναμικής φύσης των στάσεων του κοινού και μπορεί να βοηθήσει στον εντοπισμό αναδυόμενων τάσεων, ανησυχιών και λανθασμένων αντιλήψεων σχετικά με τη σπατάλη τροφίμων.

- **Πλατφόρμα για δημόσιο διάλογο:** Τα μέσα κοινωνικής δικτύωσης χρησιμεύουν ως δημόσιο φόρουμ όπου άτομα και οργανισμοί συζητούν και μοιράζονται πληροφορίες σχετικά με τη σπατάλη τροφίμων. Αυτή η ομιλία μπορεί να αποκαλύψει κοινά θέματα, κενά γνώσης και την αποτελεσματικότητα των τρεχουσών εκστρατειών ευαισθητοποίησης.
- **Επίδραση των Influencers:** Οι influencers και τα δημόσια πρόσωπα στα social media μπορούν να διαμορφώσουν σημαντικά την αντίληψη του κοινού. Οι αναρτήσεις τους σχετικά με τη σπατάλη τροφίμων μπορούν να ευαισθητοποιήσουν, να αλλάξουν νοοτροπίες και ακόμη και να εμπνεύσουν δράση στους οπαδούς τους.
- **Περιεχόμενο που δημιουργείται από τον χρήστη:** Τα μέσα κοινωνικής δικτύωσης είναι πλούσια σε περιεχόμενο που δημιουργείται από χρήστες, συμπεριλαμβανομένων αναρτήσεων, σχολίων, εικόνων και βίντεο που σχετίζονται με τα απόβλητα τροφίμων. Αυτό το περιεχόμενο παρέχει αυθεντικές πληροφορίες για το πώς οι άνθρωποι αντιλαμβάνονται και αλληλοεπιδρούν με το θέμα στην καθημερινή τους ζωή.
- **Μηχανισμός ανάδρασης:** Τα μέσα κοινωνικής δικτύωσης δίνουν τη δυνατότητα στους οργανισμούς και τους υπεύθυνους χάραξης πολιτικής να λαμβάνουν άμεση ανατροφοδότηση από το κοινό σχετικά με πρωτοβουλίες και εκστρατείες που σχετίζονται με τη σπατάλη τροφίμων. Αυτή η ανατροφοδότηση είναι πολύτιμη για τη βελτίωση των στρατηγικών και τη βελτίωση των προσπαθειών επικοινωνίας
- **Δεδομένα για Έρευνα και Ανάλυση:** Ο τεράστιος όγκος δεδομένων που δημιουργείται στα μέσα κοινωνικής δικτύωσης μπορεί να αναλυθεί χρησιμοποιώντας προηγμένες τεχνικές όπως η επεξεργασία φυσικής γλώσσας, η ανάλυση συναισθημάτων και η μοντελοποίηση θεμάτων. Αυτή η ανάλυση μπορεί να αποκαλύψει μοτίβα και ιδέες που θα ήταν δύσκολο να εντοπιστούν με άλλα μέσα.

Συνοπτικά, τα μέσα κοινωνικής δικτύωσης διαδραματίζουν κρίσιμο ρόλο στην κατανόηση της κοινής αντίληψης για τα απόβλητα τροφίμων. Προσφέρουν μια πλούσια πηγή δεδομένων που, όταν αναλυθούν αποτελεσματικά, μπορεί να παρέχει πολύτιμες πληροφορίες για τον τρόπο με τον οποίο οι άνθρωποι βλέπουν και αλληλοεπιδρούν με το ζήτημα. Αυτές οι ιδέες είναι απαραίτητες για το σχεδιασμό στοχευμένων παρεμβάσεων και στρατηγικών επικοινωνίας για την προώθηση βιώσιμων πρακτικών και τη μείωση της σπατάλης τροφίμων.

3. Τεχνολογίες Συστήματος (Technology Stack)

Σε αυτή την ενότητα παρουσιάζονται οι τεχνολογίες που χρησιμοποιήθηκαν για την ανάπτυξη του υποσυστήματος συλλογής και ανάλυσης μεγάλων δεδομένων από κοινωνικά δίκτυα.

3.1. Python - Scripting και Επεξεργασία Δεδομένων

Η Python είναι μια ευέλικτη και ευρέως χρησιμοποιούμενη γλώσσα προγραμματισμού που έπαιξε κεντρικό ρόλο σε αυτό το αναπτυχθέν σύστημα, ιδιαίτερα στη δέσμη ενεργειών και την επεξεργασία δεδομένων. Η σημασία και οι εφαρμογές του στο αναπτυχθέν σύστημα περιγράφονται παρακάτω:

- **Ευκολία χρήσης και αναγνωσιμότητα:** Η απλή σύνταξη και η αναγνωσιμότητα της Python την καθιστούν ιδανική γλώσσα για δέσμες ενεργειών και επεξεργασία δεδομένων. Η απλή δομή της επιτρέπει τη γρήγορη ανάπτυξη και την εύκολη συντήρηση του κώδικα.
- **Πλούσιο Οικοσύστημα Βιβλιοθήκης:** Η Python διαθέτει μια τεράστια συλλογή βιβλιοθηκών και πλαισίων που είναι ιδιαίτερα κατάλληλα για εργασίες επεξεργασίας δεδομένων. Βιβλιοθήκες όπως οι Pandas για χειρισμό δεδομένων, NumPy για αριθμητικούς υπολογισμούς και Matplotlib για οπτικοποίηση δεδομένων χρησιμοποιούνται εκτενώς στο αναπτυχθέν σύστημα.
- **Επεξεργασία φυσικής γλώσσας (NLP):** Οι βιβλιοθήκες NLP της Python, όπως το NLTK και το spaCy, είναι καθοριστικής σημασίας για την επεξεργασία και την ανάλυση του κειμένου των μέσων κοινωνικής δικτύωσης. Αυτές οι βιβλιοθήκες παρέχουν λειτουργίες για διανυσματοποίηση και λήμματοποίηση και ανάλυση συναισθήματος, οι οποίες είναι ζωτικής σημασίας για την κατανόηση του περιεχομένου και του συναισθήματος των αναρτήσεων στα μέσα κοινωνικής δικτύωσης.
- **Συλλογή δεδομένων και ενσωμάτωση:** Τα σενάρια Python χρησιμοποιούνται για διασύνδεση με διάφορα API, όπως το Twitter API, για τη συλλογή δεδομένων μέσω κοινωνικής δικτύωσης. Η ικανότητα της Python να χειρίζεται διαφορετικές μορφές δεδομένων και πηγές απλοποιεί τη διαδικασία ενσωμάτωσης δεδομένων από πολλές πλατφόρμες όπως το Twitter και το Reddit.
- **Μηχανική μάθηση και ανάλυση:** Για προηγμένη ανάλυση δεδομένων, χρησιμοποιούνται βιβλιοθήκες μηχανικής μάθησης της Python, συμπεριλαμβανομένων των scikit-learn και TensorFlow. Αυτές οι βιβλιοθήκες διευκολύνουν την εφαρμογή αλγορίθμων για μοντελοποίηση θεμάτων, ανάλυση συναισθήματος και προγνωστική ανάλυση.
- **Επεκτασιμότητα και απόδοση:** Η συμβατότητα της Python με άλλες τεχνολογίες και η ικανότητά της να ενσωματώνεται/συνεργάζεται με εργαλεία μεγάλων δεδομένων όπως το Hadoop /Apache Spark και τις NoSQL ΒΔ, την καθιστούν μια επεκτάσιμη επιλογή για την επεξεργασία μεγάλου όγκου δεδομένων μέσω κοινωνικής δικτύωσης.
- **Αυτοματισμός σεναρίων:** Τα σενάρια Python χρησιμοποιούνται σε συνδυασμό με εργαλεία όπως το cron για την αυτοματοποίηση των εργασιών συλλογής και επεξεργασίας

δεδομένων. Αυτός ο αυτοματισμός διασφαλίζει ότι τα πιο πρόσφατα δεδομένα συλλέγονται και επεξεργάζονται συνεχώς χωρίς χειροκίνητη παρέμβαση.

- **Υποστήριξη κοινότητας:** Η Python έχει μια μεγάλη και ενεργή κοινότητα, η οποία παρέχει εκτεταμένη υποστήριξη μέσω τεκμηρίωσης, φόρουμ και πόρων τρίτων. Αυτή η υποστήριξη της κοινότητας είναι ανεκτίμητη για την αντιμετώπιση προβλημάτων και τη βελτίωση των δυνατοτήτων του αναπτυχθέντος συστήματος.

Συμπερασματικά, η ευελιξία της Python, σε συνδυασμό με τις ισχυρές βιβλιοθήκες και την υποστήριξη της κοινότητας, την καθιστούν ιδανική επιλογή για τις ανάγκες σεναρίου και επεξεργασίας δεδομένων του συστήματος. Οι δυνατότητές του επιτρέπουν τον αποτελεσματικό χειρισμό δεδομένων κοινωνικής δικτύωσης μεγάλης κλίμακας, από τη συλλογή έως την ανάλυση, διαδραματίζοντας καθοριστικό ρόλο στην απόκτηση ουσιαστικών γνώσεων από τα δεδομένα.

3.2. Twitter API και Reddit API - Συλλογή δεδομένων

Το Twitter API και το Reddit API αποτελούν αναπόσπαστα στοιχεία του συστήματος, επιτρέποντας τη συστηματική συλλογή δεδομένων μέσω κοινωνικής δικτύωσης. Οι ρόλοι και οι λειτουργίες τους στο πλαίσιο του αναπτυχθέντος συστήματος είναι οι εξής:



1. Πρόσβαση σε δεδομένα μέσω κοινωνικής δικτύωσης: Και τα δύο API παρέχουν πρόσβαση μέσω προγραμματισμού σε τεράστιες ποσότητες περιεχομένου κοινωνικών μέσων. Το Twitter API επιτρέπει την ανάκτηση tweet, προφίλ χρηστών και σχετικών μεταδεδομένων, ενώ το Reddit API διευκολύνει την πρόσβαση σε αναρτήσεις, σχόλια και πληροφορίες subreddit.

2. Προσαρμόσιμα ερωτήματα δεδομένων: Και τα δύο API προσφέρουν την ευελιξία για την προσαρμογή των ερωτημάτων με βάση συγκεκριμένες λέξεις-κλειδιά, hashtags, αναγνωριστικά χρήστη και άλλες παραμέτρους. Αυτή η δυνατότητα είναι ζωτικής σημασίας για τη στόχευση περιεχομένου που σχετίζεται με τα απορρίμματα τροφίμων και για τη διασφάλιση της συνάφειας των δεδομένων που συλλέγονται.

3. Συλλογή δεδομένων σε πραγματικό χρόνο: Και τα δύο API επιτρέπουν τη συλλογή δεδομένων σε πραγματικό χρόνο, κάτι που είναι απαραίτητο για την καταγραφή των τρεχουσών τάσεων, των συζητήσεων και των απόψεων της κοινής γνώμης σχετικά με τη σπατάλη τροφίμων. Αυτή η έγκαιρη συλλογή δεδομένων βοηθά στη διατήρηση ενός ενημερωμένου δεδομένων για ανάλυση.

4. Πρόσβαση σε ιστορικά δεδομένα: Εκτός από δεδομένα σε πραγματικό χρόνο, και τα δύο API παρέχουν επίσης πρόσβαση σε ιστορικά δεδομένα, επιτρέποντας διαχρονικές μελέτες και ανάλυση τάσεων με την πάροδο του χρόνου. Αυτή η ιστορική προοπτική είναι πολύτιμη για την κατανόηση της εξέλιξης του δημόσιου λόγου σχετικά με τα απόβλητα τροφίμων.

5. Διαχείριση ορίων ποσοστού: Τόσο τα API του Twitter όσο και του Reddit έχουν όρια ρυθμού για τον έλεγχο του όγκου των δεδομένων στα οποία είναι δυνατή η πρόσβαση σε ένα δεδομένο χρονικό πλαίσιο. Ο αποτελεσματικός χειρισμός αυτών των ορίων ρυθμού μέσω της δέσμης ενεργειών Python διασφαλίζει τη συνεχή και αδιάλειπτη συλλογή δεδομένων.

6. Δομή και Μορφή Δεδομένων: Τα API επιστρέφουν δεδομένα σε δομημένες μορφές όπως το JSON, τα οποία αναλύονται και επεξεργάζονται εύκολα χρησιμοποιώντας Python. Αυτή η δομημένη μορφή απλοποιεί την εξαγωγή σχετικών πληροφοριών, όπως περιεχόμενο κειμένου, χρονικές σημάνσεις, πληροφορίες χρήστη και μετρήσεις αφοσίωσης.



7. Ενσωμάτωση με Εργαλεία Επεξεργασίας Δεδομένων: Τα δεδομένα που συλλέγονται μέσω αυτών των API ενσωματώνονται άψογα με εργαλεία επεξεργασίας δεδομένων που βασίζονται σε Python για περαιτέρω ανάλυση. Αυτή η ενοποίηση είναι ζωτικής σημασίας για την αυτοματοποίηση της ροής εργασιών από τη συλλογή δεδομένων έως την ανάλυση.

8. Συμμόρφωση με τις Πολιτικές Πλατφόρμας: Η χρήση επίσημων API διασφαλίζει τη συμμόρφωση με τις πολιτικές και τους όρους παροχής υπηρεσιών της αντίστοιχης πλατφόρμας, κάτι που είναι σημαντικό για τις ηθικές πρακτικές συλλογής δεδομένων.

Συνοπτικά, το Twitter API και το Reddit API είναι απαραίτητα για την αποτελεσματική συλλογή σχετικών δεδομένων κοινωνικών μέσων για το έργο. Η ικανότητά τους να παρέχουν δεδομένα σε πραγματικό χρόνο και ιστορικά δεδομένα, σε συνδυασμό με προσαρμόσιμες επιλογές ερωτημάτων και δομημένες μορφές δεδομένων, τα καθιστά ανεκτίμητα εργαλεία για τη συλλογή πληροφοριών σχετικά με τις αντιλήψεις του κοινού για τα απόβλητα τροφίμων. Η ενοποίηση αυτών των API με σενάρια Python ενισχύει περαιτέρω την αποδοτικότητα και την αποτελεσματικότητα της διαδικασίας συλλογής δεδομένων.

3.3. MongoDB - Αποθήκευση Δεδομένων

Η MongoDB, μια βάση δεδομένων NoSQL, διαδραματίζει σημαντικό ρόλο στο αναπτυχθέν σύστημα ως η κύρια λύση αποθήκευσης για τα δεδομένα κοινωνικών μέσων που συλλέγονται. Τα χαρακτηριστικά και τα οφέλη του στο πλαίσιο αυτού του αναπτυχθέντος συστήματος περιλαμβάνουν:

- **Μοντέλο δεδομένων χωρίς σχήμα:** Η δυνατότητα της MongoDB για τη διαχείριση δεδομένων χωρίς σχήμα επιτρέπει την αποθήκευση μη δομημένων και ημιδομημένων δεδομένων, κάτι που είναι χαρακτηριστικό στο περιεχόμενο των μέσων κοινωνικής δικτύωσης. Αυτή η ευελιξία είναι ιδιαίτερα χρήσιμη για την προσαρμογή των ποικίλων μορφών και δομών δεδομένων που συναντώνται στα δεδομένα Twitter και Reddit και αυξάνει την αποδοτικότητα των ερωτημάτων

- **Επεκτασιμότητα:** Η MongoDB υποστηρίζει την υλοποίηση εξαιρετικά επεκτάσιμων συστημάτων (scaleup/scaledown), γεγονός που τη καθιστά κατάλληλη για το χειρισμό μεγάλων όγκων δεδομένων. Αυτή η επεκτασιμότητα διασφαλίζει ότι η βάση δεδομένων μπορεί να αναπτυχθεί ανάλογα με τις ανάγκες του αναπτυσσόμενου συστήματος, φιλοξενώντας αυξανόμενες ποσότητες δεδομένων μέσω κοινωνικής δικτύωσης με την πάροδο του χρόνου.
- **Υψηλή απόδοση:** Η MongoDB προσφέρει υψηλή απόδοση, ειδικά όσον αφορά τις λειτουργίες ανάγνωσης και εγγραφής, κάτι που είναι απαραίτητο για την επεξεργασία και ανάλυση δεδομένων σε πραγματικό χρόνο. Οι αποτελεσματικές δυνατότητες δημιουργίας ευρετηρίου και αναζήτησης δεδομένων επιτρέπουν τη γρήγορη ανάκτηση και χειρισμό μεγάλων συνόλων δεδομένων. Οι προσφερόμενες δυνατότητες για ομοιοτυπία (replication) και τμηματοποίηση (sharding) αυξάνουν την απόδοση, διαθεσιμότητα (availability) και την αξιοπιστία των συστημάτων MongoDB. Επιπλέον υποστηρίζει πραγματικά κατανεμημένη λειτουργία (distributed queries, distributed mongos) αυξάνοντας σημαντικά την απόδοση αλλά και την επεκτασιμότητα της ΒΔ.
- **Έγγραφα JSON:** Τα δεδομένα στη MongoDB αποθηκεύονται σε μορφή BSON (Binary JSON), η οποία είναι παρόμοια με JSON. Αυτή η μορφή ευθυγραμμίζεται καλά με τις δομές δεδομένων που χρησιμοποιούνται στην Python και στα API, απλοποιώντας τη διαδικασία εισαγωγής, ανάκτησης και επεξεργασίας δεδομένων.
- **Υποστήριξη για σύνθετα ερωτήματα:** Η MongoDB υποστηρίζει σύνθετα ερωτήματα, συμπεριλαμβανομένων των αναζητήσεων κειμένου και των λειτουργιών συγκέντρωσης (aggregation), που είναι ζωτικής σημασίας για την ανάλυση δεδομένων μέσω κοινωνικής δικτύωσης. Αυτές οι δυνατότητες επιτρέπουν την εξελιγμένη ανάλυση δεδομένων απευθείας μέσα στη βάση δεδομένων. Επίσης, το ισχυρό πλαίσιο συγκέντρωσης της MongoDB επιτρέπει την εκτέλεση πολύπλοκων λειτουργιών επεξεργασίας και μετασχηματισμού δεδομένων εντός της βάσης δεδομένων. Αυτή η δυνατότητα είναι ιδιαίτερα χρήσιμη για την προεπεξεργασία και την ανάλυση μεγάλων συνόλων δεδομένων χωρίς την ανάγκη πρόσθετων εργαλείων επεξεργασίας δεδομένων.
- **Στιβαρότητα και αξιοπιστία:** Η MongoDB παρέχει στιβαρότητα και αξιοπιστία με λειτουργίες όπως η αναπαραγωγή και η αυτόματη ανακατεύθυνση, διασφαλίζοντας την ακεραιότητα και τη διαθεσιμότητα των δεδομένων ακόμη και σε περίπτωση αστοχίας του συστήματος.
- **Ενσωμάτωση με Εργαλεία Ανάλυσης Δεδομένων:** Η MongoDB ενσωματώνεται απρόσκοπτα με την Python και διάφορα εργαλεία ανάλυσης δεδομένων, διευκολύνοντας την ομαλή ροή εργασίας από την αποθήκευση δεδομένων στην ανάλυση. Αυτή η ενοποίηση είναι το κλειδί για την αποτελεσματική επεξεργασία και οπτικοποίηση των δεδομένων των μέσω κοινωνικής δικτύωσης.



Συμπερασματικά, η ευέλικτη και επεκτάσιμη αρχιτεκτονική του MongoDB, σε συνδυασμό με τις ισχυρές δυνατότητες αναζήτησης και συγκέντρωσης, το καθιστά ιδανική επιλογή για την αποθήκευση και τη διαχείριση των ποικίλων και ογκωδών δεδομένων κοινωνικών μέσων που συλλέγονται σε αυτό το έργο. Η συμβατότητά του με την Python και άλλα εργαλεία ανάλυσης δεδομένων ενισχύει περαιτέρω τη χρησιμότητά του, επιτρέποντας την αποτελεσματική διαχείριση και ανάλυση δεδομένων.

3.4. Dash Plotly - Δημιουργία Dashboard

Το Dash Plotly είναι μια κρίσιμη τεχνολογία, που χρησιμοποιείται για τη δημιουργία διαδραστικών και δυναμικών πινάκων εργαλείων. Ο ρόλος και τα πλεονεκτήματά του στο πλαίσιο αυτού του αναπτυχθέντος συστήματος περιγράφονται παρακάτω:

1. Διαδραστική Οπτικοποίηση Δεδομένων: Το Dash Plotly υπερέχει στη δημιουργία διαδραστικών οπτικοποιήσεων δεδομένων, οι οποίες είναι απαραίτητες για την αποτελεσματική επικοινωνία περίπλοκων πληροφοριών δεδομένων. Οι χρήστες μπορούν να αλληλεπιδράσουν με τις απεικονίσεις, όπως φιλτράρισμα δεδομένων, μεγέθυνση σε λεπτομέρειες και προβολή συγκεκριμένων μετρήσεων, βελτιώνοντας την κατανόησή τους για τα δεδομένα.

2. Ενσωμάτωση με την Python: Το Dash είναι χτισμένο πάνω από το Plotly και είναι πλήρως ενσωματωμένο με την Python, επιτρέποντας την απρόσκοπτη χρήση των εκτεταμένων δυνατοτήτων επεξεργασίας και ανάλυσης δεδομένων της Python. Αυτή η ενοποίηση είναι ιδιαίτερα επωφελής για το αναπτυχθέν σύστημα, καθώς η Python χρησιμοποιείται για συλλογή και επεξεργασία δεδομένων.

3. Προσαρμόσιμες διατάξεις και στοιχεία: Το Dash προσφέρει ένα ευρύ φάσμα προσαρμόσιμων διατάξεων και στοιχείων, επιτρέποντας τη δημιουργία προσαρμοσμένων πινάκων εργαλείων που πληρούν συγκεκριμένες απαιτήσεις έργου. Αυτή η ευελιξία επιτρέπει το σχεδιασμό πινάκων εργαλείων που παρουσιάζουν αποτελεσματικά τις γνώσεις που προέρχονται από δεδομένα μέσω κοινωνικής δικτύωσης που σχετίζονται με τα απόβλητα τροφίμων.

4. Ενημερώσεις δεδομένων σε πραγματικό χρόνο: Το Dash υποστηρίζει ενημερώσεις δεδομένων σε πραγματικό χρόνο, κάτι που είναι ζωτικής σημασίας για την εμφάνιση των πιο πρόσφατων δεδομένων και πληροφοριών. Αυτή η δυνατότητα είναι ιδιαίτερα σημαντική για έργα που περιλαμβάνουν συνεχή συλλογή και ανάλυση δεδομένων, όπως η παρακολούθηση των τάσεων των μέσων κοινωνικής δικτύωσης.

5. Ευκολία ανάπτυξης: Οι εφαρμογές Dash μπορούν εύκολα να αναπτυχθούν σε διάφορες πλατφόρμες, καθιστώντας τους πίνακες εργαλείων προσβάσιμους σε ένα ευρύτερο κοινό. Αυτή η ευκολία ανάπτυξης είναι ζωτικής σημασίας για τη διάδοση των πορισμάτων του έργου στα ενδιαφερόμενα μέρη και στο κοινό.

6. Σχεδίαση με απόκριση: Οι πίνακες εργαλείων Dash Plotly έχουν απόκριση, δηλαδή προσαρμόζουν τη διάταξη και τη λειτουργικότητά τους με βάση τη συσκευή στην οποία προβάλλονται (π.χ. επιτραπέζιος υπολογιστής, tablet, smartphone). Αυτή η ανταπόκριση εξασφαλίζει μια συνεπή εμπειρία χρήστη σε διαφορετικές συσκευές.

7. Εκτεταμένη τεκμηρίωση και υποστήριξη κοινότητας: Το Dash Plotly διαθέτει εκτενή τεκμηρίωση και μια υποστηρικτική κοινότητα, η οποία είναι επωφελής για την αντιμετώπιση προβλημάτων, την εκμάθηση βέλτιστων πρακτικών και την εξερεύνηση προηγμένων λειτουργιών. Αυτή η υποστήριξη είναι ανεκτίμητη για την αποτελεσματική ανάπτυξη και συντήρηση των dashboard του αναπτυχθέντος συστήματος.

8. Λήψη αποφάσεων βάσει δεδομένων: Οπτικοποιώντας τα επεξεργασμένα δεδομένα μέσω κοινωνικής δικτύωσης σε προσβάσιμη και κατανοητή μορφή, οι πίνακες εργαλείων Dash Plotly διευκολύνουν τη λήψη αποφάσεων βάσει δεδομένων. Τα ενδιαφερόμενα μέρη μπορούν να χρησιμοποιήσουν αυτούς τους πίνακες εργαλείων για να αντλήσουν πληροφορίες που να μπορούν να κάνουν πράξη και να λάβουν τεκμηριωμένες αποφάσεις σχετικά με πρωτοβουλίες για τα απόβλητα τροφίμων.

Συνοπτικά, οι δυνατότητες του Dash Plotly στη δημιουργία διαδραστικών, προσαρμόσιμων και εύχρηστων πινάκων εργαλείων το καθιστούν ιδανική επιλογή για αυτό το έργο. Η ενσωμάτωση του με την Python, σε συνδυασμό με τη δυνατότητα ενημέρωσης δεδομένων σε πραγματικό χρόνο και τον αποκριτικό σχεδιασμό, διασφαλίζει ότι οι πίνακες εργαλείων μεταφέρουν αποτελεσματικά τις γνώσεις που προκύπτουν από την ανάλυση των δεδομένων των μέσων κοινωνικής δικτύωσης για τα απόβλητα τροφίμων.

3.5. Άλλες Βιβλιοθήκες και Εργαλεία

Εκτός από τις κύριες τεχνολογίες, το έργο αξιοποιεί διάφορες βιβλιοθήκες και εργαλεία για εξειδικευμένες εργασίες, ιδιαίτερα στην Επεξεργασία Φυσικής Γλώσσας (NLP) και στην οπτικοποίηση δεδομένων. Βασικά μεταξύ αυτών είναι τα NLTK, Gensim και PyLDAvis.

3.5.1. NLTK (Εργαλειοθήκη φυσικής γλώσσας)

Το NLTK είναι μια κορυφαία πλατφόρμα για τη δημιουργία προγραμμάτων Python για εργασία με δεδομένα ανθρώπινης γλώσσας. Παρέχει εύχρηστες διεπαφές σε περισσότερα από 50 σώματα και λεξιλογικούς πόρους. Το NLTK χρησιμοποιείται για διάφορες εργασίες NLP, όπως την τμηματοποίηση κειμένου (tokenization), την απομάκρυνση κοινών λέξεων (stop words removal) την θεματοποίηση (stemming) και τη λημματοποίηση (lemmatization). Αυτές οι διαδικασίες είναι απαραίτητες για την προεπεξεργασία και την κανονικοποίηση των δεδομένων κειμένου των μέσων κοινωνικής δικτύωσης με σκοπό την περαιτέρω ανάλυση.

Η εκτεταμένη βιβλιοθήκη πόρων και εργαλείων του NLTK το καθιστά εξαιρετικά αποτελεσματικό για την επεξεργασία γλωσσικών δεδομένων. Είναι ιδιαίτερα χρήσιμο για τον

χειρισμό της πολυπλοκότητας και των αποχρώσεων της φυσικής γλώσσας στο περιεχόμενο των μέσων κοινωνικής δικτύωσης.

3.5.2. **GenSim**

Το Gensim είναι μια βιβλιοθήκη ανοιχτού κώδικα σχεδιασμένη για μοντελοποίηση θεμάτων χωρίς επίβλεψη και επεξεργασία φυσικής γλώσσας. Είναι ιδιαίτερα γνωστό για την εφαρμογή του Word2Vec, LDA (Latent Dirichlet Allocation) και άλλων αλγορίθμων μοντελοποίησης θεμάτων (Topics Modeling and Topics Extraction). Το Gensim χρησιμοποιείται για εργασίες μοντελοποίησης θεμάτων για την εξαγωγή θεμάτων από τα δεδομένα κοινωνικών μέσων που συλλέγονται. Αυτό βοηθά στην κατανόηση των διαδεδομένων συζητήσεων και των συναισθημάτων γύρω από τη σπατάλη τροφίμων. Οι αποτελεσματικές και επεκτάσιμες υλοποιήσεις αλγορίθμων μοντελοποίησης θεμάτων του Gensim είναι ζωτικής σημασίας για το χειρισμό μεγάλων συνόλων δεδομένων, τυπικών στην ανάλυση των μέσων κοινωνικής δικτύωσης. Η ικανότητά του να αποκαλύπτει υποκείμενες σημασιολογικές δομές σε δεδομένα κειμένου προσθέτει βάθος στην ανάλυση.

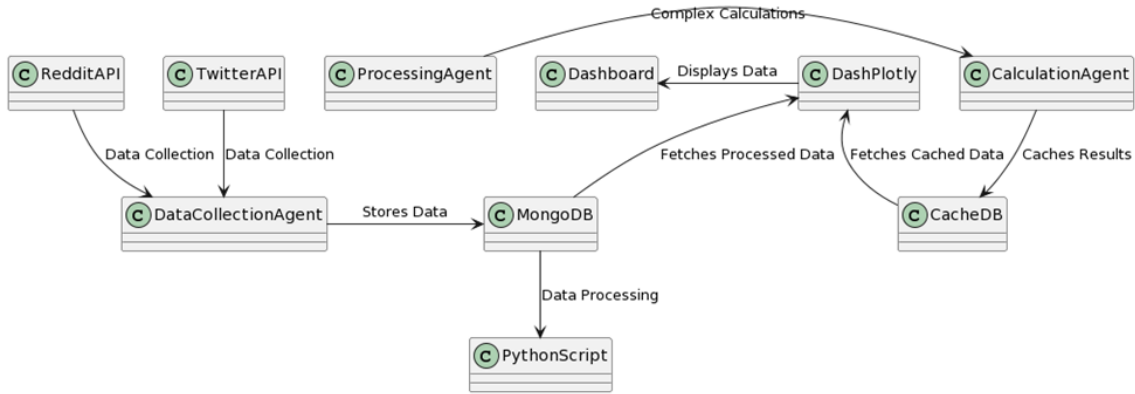
3.5.3. **PyLDAvis**

Το PyLDAvis είναι μια βιβλιοθήκη Python για διαδραστική οπτικοποίηση μοντέλων θεμάτων. Έχει σχεδιαστεί για να βοηθά τους χρήστες να ερμηνεύουν τα θέματα σε ένα μοντέλο θέματος. Το PyLDAvis χρησιμοποιείται σε συνδυασμό με το Gensim για την οπτικοποίηση των αποτελεσμάτων της μοντελοποίησης θεμάτων. Παρέχει μια διαδραστική οπτικοποίηση που βασίζεται στον ιστό που βοηθά στη διερεύνηση των σχέσεων μεταξύ των θεμάτων και της κατανομής των λέξεων σε κάθε θέμα. Το PyLDAvis ενισχύει την ερμηνευτικότητα των μοντέλων θεμάτων παρέχοντας μια σαφή και διαδραστική οπτική αναπαράσταση. Αυτό διευκολύνει τους ενδιαφερόμενους να κατανοήσουν τα αποτελέσματα μοντελοποίησης περίπλοκων θεμάτων και να αντλήσουν σημαντικές γνώσεις.

Αυτές οι βιβλιοθήκες και τα εργαλεία διαδραματίζουν σημαντικό ρόλο στο έργο παρέχοντας προηγμένες δυνατότητες επεξεργασίας και ανάλυσης δεδομένων φυσικής γλώσσας. Η ενσωμάτωσή τους στη ροή εργασίας του έργου επιτρέπει μια ολοκληρωμένη ανάλυση του περιεχομένου των μέσων κοινωνικής δικτύωσης, οδηγώντας σε βαθύτερες γνώσεις σχετικά με τις αντιλήψεις του κοινού και τις συζητήσεις σχετικά με τα απορρίμματα τροφίμων.

4. Αρχιτεκτονική Συστήματος

Η αρχιτεκτονική του συστήματος για το έργο έχει σχεδιαστεί για να χειρίζεται αποτελεσματικά τη συλλογή, την επεξεργασία, την ανάλυση και την οπτικοποίηση δεδομένων μέσω κοινωνικής δικτύωσης που σχετίζονται με τα απόβλητα τροφίμων. Ακολουθεί ένα εννοιολογικό διάγραμμα της αρχιτεκτονικής του συστήματος, στο **Σχήμα 4.1**:



Σχήμα 4-1: Αρχιτεκτονική Συστήματος (Conceptual Diagram)

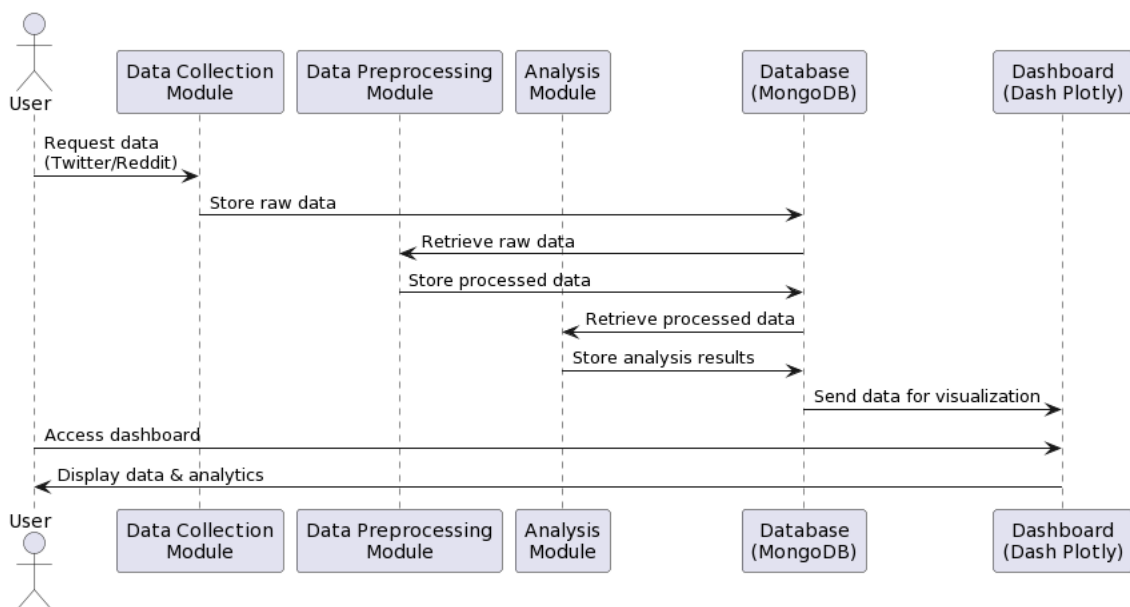
Λειτουργικότητα		Τεχνολογία	Λεπτομέρειες
Επίπεδο συλλογής δεδομένων	συλλογής	Twitter API, Reddit API	Αυτό το επίπεδο είναι υπεύθυνο για τη συλλογή δεδομένων από το Twitter και το Reddit. Τα API χρησιμοποιούνται για την αναζήτηση και την ανάκτηση αναρτήσεων, σχολίων και σχετικών μεταδεδομένων με βάση καθορισμένες λέξεις-κλειδιά και κριτήρια.
Επίπεδο διαχείρισης/αποθήκευσης δεδομένων		MongoDB	Λειτουργία: Τα δεδομένα που συλλέγονται αποθηκεύονται στο MongoDB, μια βάση δεδομένων NoSQL, η οποία παρέχει ευελιξία και επεκτασιμότητα για το χειρισμό μεγάλων όγκων μη δομημένων δεδομένων. Το MongoDB χρησιμοποιείται για την αποθήκευση ακατέργαστων δεδομένων από πλατφόρμες μέσω κοινωνικής δικτύωσης καθώς και επεξεργασμένα και αναλυμένα δεδομένα.

Επίπεδο επεξεργασίας και ανάλυσης δεδομένων	Σενάρια Python, εργαλεία NLP (NLTK, Gensim).	Αυτό το επίπεδο περιλαμβάνει την προεπεξεργασία ακατέργαστων δεδομένων (π.χ. κανονικοποίηση κειμένου, tokenization, stemming, lemmatization) και την ανάλυσή τους. Η ανάλυση συναισθήματος και η μοντελοποίηση θεμάτων εκτελούνται χρησιμοποιώντας εργαλεία NLP για την εξαγωγή πληροφοριών από τα δεδομένα.
Οπτικοποίηση δεδομένων και Επίπεδο πίνακα ελέγχου:	Dash Plotly	Τα επεξεργασμένα και αναλυμένα δεδομένα οπτικοποιούνται χρησιμοποιώντας το Dash Plotly, το οποίο παρέχει έναν διαδραστικό πίνακα εργαλείων που βασίζεται στον ιστό. Ο πίνακας εργαλείων επιτρέπει στους χρήστες να εξερευνούν διάφορες πτυχές των δεδομένων, όπως τάσεις συναισθήματος, συχνότητες λέξεων και διανομές θεμάτων.
Επίπεδο προσωρινής αποθήκευσης και βελτιστοποίησης απόδοσης	MongoDB (για προσωρινή αποθήκευση), σενάρια Python	Για να βελτιωθεί η απόδοση του πίνακα εργαλείων, ειδικά για εργασίες με υπολογιστική ένταση, όπως η δημιουργία γραφημάτων ταυτόχρονης εμφάνισης, τα αποτελέσματα αποθηκεύονται προσωρινά στο MongoDB. Αυτό μειώνει τον χρόνο φόρτωσης και βελτιώνει την εμπειρία του χρήστη.
Επίπεδο προγραμματισμού και αυτοματισμού:	Cron jobs.	Τα αυτοματοποιημένα σενάρια προγραμματίζονται με τη χρήση εργασιών cron για την περιοδική συλλογή δεδομένων από πλατφόρμες μέσων κοινωνικής δικτύωσης και την εκτέλεση εργασιών ρουτίνας επεξεργασίας και ανάλυσης δεδομένων.

Η αρχιτεκτονική έχει σχεδιαστεί ώστε να είναι αρθρωτή και επεκτάσιμη, επιτρέποντας την ενσωμάτωση πρόσθετων πηγών δεδομένων, αναλυτικών εργαλείων ή στοιχείων οπτικοποίησης, όπως απαιτείται. Η χρήση σύγχρονων τεχνολογιών και πρακτικών διασφαλίζει ότι το σύστημα μπορεί να χειριστεί μεγάλης κλίμακας επεξεργασία δεδομένων και να παρέχει πληροφορίες σε πραγματικό χρόνο για τις αντιλήψεις του κοινού για τα απόβλητα τροφίμων.

Στο επόμενο **Σχήμα 4.2**, δίνεται το διάγραμμα αλληλουχίας (Sequence Diagram) του συστήματος σε σημειολογία της Unified Modelling Language – UML notation. Αυτό το διάγραμμα περιγράφει τα ακόλουθα βήματα:

- Ο χρήστης ζητά δεδομένα από τη μονάδα συλλογής δεδομένων.
- Το Data Collection Module συλλέγει δεδομένα από το Twitter και το Reddit και τα αποθηκεύει στη βάση δεδομένων MongoDB.
- Η μονάδα προεπεξεργασίας δεδομένων ανακτά ανεπεξέργαστα δεδομένα από τη βάση δεδομένων, τα επεξεργάζεται και αποθηκεύει τα επεξεργασμένα δεδομένα πίσω στη βάση δεδομένων.
- Η μονάδα ανάλυσης ανακτά επεξεργασμένα δεδομένα, εκτελεί αναλύσεις (όπως ανάλυση συναισθήματος, ανάλυση συχνότητας, μοντελοποίηση θεμάτων) και αποθηκεύει τα αποτελέσματα στη βάση δεδομένων.
- Ο πίνακας εργαλείων (κατασκευασμένος με το Dash Plotly) ανακτά δεδομένα και αποτελέσματα ανάλυσης από τη βάση δεδομένων και τα εμφανίζει στον χρήστη.

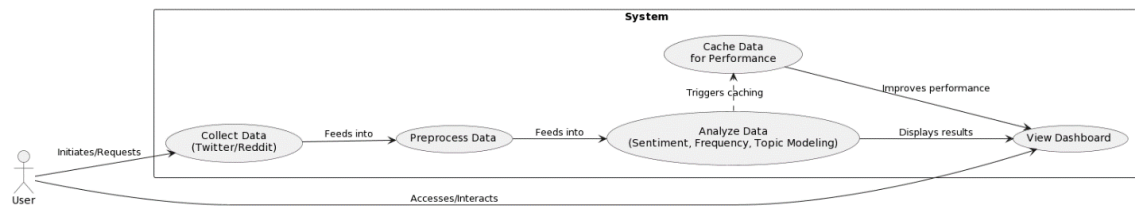


Σχήμα 4-2: Διάγραμμα UML - Sequence

Όμοια στο **Σχήμα 4.3**, δίνεται το διάγραμμα αλληλεπιδράσεων του συστήματος, που παρουσιάζει τις ακόλουθες αλληλεπιδράσεις:

- Ο χρήστης ξεκινά ή ζητά συλλογή δεδομένων από πλατφόρμες μέσων κοινωνικής δικτύωσης (Twitter/Reddit).

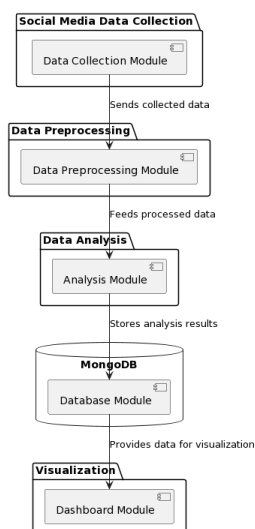
- Στη συνέχεια, τα δεδομένα που συλλέγονται υποβάλλονται σε προεπεξεργασία.
- Τα προεπεξεργασμένα δεδομένα υποβάλλονται σε διάφορες αναλύσεις, συμπεριλαμβανομένης της ανάλυσης συναισθήματος, της ανάλυσης συχνότητας και της μοντελοποίησης θεμάτων.
- Τα αποτελέσματα αυτών των αναλύσεων εμφανίζονται σε έναν πίνακα εργαλείων, τον οποίο ο χρήστης μπορεί να δει και να αλληλεπιδράσει μαζί του.
- Το σύστημα περιλαμβάνει επίσης έναν μηχανισμό προσωρινής αποθήκευσης για τη βελτίωση της απόδοσης του dashboard.



Σχήμα 4-3: Διάγραμμα Αλληλοεπιδράσεων (interactions)

Τέλος, στο **Σχήμα 4.4**, δίνεται το διάγραμμα αλληλοεπιδράσεων ανάμεσα στα components του συστήματος. Πιο συγκεκριμένα, αυτό το σενάριο περιγράφει τις ακόλουθες αλληλεπιδράσεις στοιχείων:

- Η μονάδα συλλογής δεδομένων συλλέγει δεδομένα από πλατφόρμες μέσων κοινωνικής δικτύωσης και τα στέλνει στη μονάδα προεπεξεργασίας δεδομένων.
- Η μονάδα προεπεξεργασίας δεδομένων επεξεργάζεται τα δεδομένα και τα τροφοδοτεί στη μονάδα ανάλυσης.
- Η μονάδα ανάλυσης εκτελεί διάφορες αναλύσεις στα δεδομένα και αποθηκεύει τα αποτελέσματα στη μονάδα βάσης δεδομένων.
- Η Μονάδα Βάσης Δεδομένων, που υλοποιείται με τη χρήση MongoDB, παρέχει τα απαραίτητα δεδομένα για οπτικοποίηση στη μονάδα του πίνακα ελέγχου.



Σχήμα 4-4: Διάγραμμα Components Interaction

4.1. Docker Deployment

Στο πλαίσιο της ανάπτυξης αυτού του συστήματος, ο Docker χρησιμοποιείται για την αποθήκευση διαφόρων στοιχείων του συστήματος ανάλυσης μέσω κοινωνικής δικτύωσης. Τα containers Docker παρέχουν ένα απομονωμένο, συνεπές και φορητό περιβάλλον για κάθε υπηρεσία, διασφαλίζοντας ότι η εφαρμογή εκτελείται ομαλά σε διαφορετικά περιβάλλοντα υπολογιστών.

Components

- **Υπηρεσία συλλογής δεδομένων:** Αυτό το container εκτελεί τα σενάρια που αλληλοεπιδρούν με τα API του Twitter και του Reddit για τη συλλογή δεδομένων. Με την αποθήκευση αυτής της υπηρεσίας, διασφαλίζουμε τη συνεπή συλλογή δεδομένων ανεξάρτητα από το περιβάλλον του κεντρικού υπολογιστή.
- **Υπηρεσία Προεπεξεργασίας Δεδομένων:** Αυτή η υπηρεσία, που εκτελείται στο δικό της container, χειρίζεται την προεπεξεργασία των συλλεγόμενων δεδομένων, συμπεριλαμβανομένης της κανονικοποίησης κειμένου, της αφαίρεσης κοινών λέξεων και των υπόλοιπων λειτουργιών προεπεξεργασίας κειμένων. Το Containerization εγγυάται ότι τα βήματα προεπεξεργασίας εκτελούνται σε ομοιόμορφο περιβάλλον, αποφεύγοντας τις αποκλίσεις λόγω διαφορετικών τοπικών ρυθμίσεων.
- **Υπηρεσία ανάλυσης:** Το στοιχείο ανάλυσης, το οποίο περιλαμβάνει ανάλυση συναισθήματος, ανάλυση συχνότητας και μοντελοποίηση θεμάτων, είναι επίσης container. Αυτή η απομόνωση διασφαλίζει ότι το υπολογιστικό περιβάλλον για την ανάλυση είναι σταθερό και αναπαραγώγιμο.
- **Υπηρεσία βάσης δεδομένων:** Η MongoDB εκτελείται σε container Docker, παρέχοντας ένα συνεπές και απομονωμένο περιβάλλον βάσης δεδομένων. Αυτή η ρύθμιση απλοποιεί τη διαχείριση και την κλιμάκωση της βάσης δεδομένων, διασφαλίζοντας ότι η βάση δεδομένων συμπεριφέρεται με τον ίδιο τρόπο σε διαφορετικές αναπτύξεις.
- **Υπηρεσία πίνακα εργαλείων:** Ο πίνακας εργαλείων Dash Plotly φιλοξενείται σε ξεχωριστό container. Αυτή η προσέγγιση επιτρέπει εύκολες ενημερώσεις και κλιμάκωση του πίνακα εργαλείων χωρίς να επηρεάζονται άλλες υπηρεσίες.

Πλεονεκτήματα

- **Συνέπεια:** Το Docker διασφαλίζει ότι κάθε υπηρεσία εκτελείται σε ίδιο περιβάλλον, μειώνοντας το πρόβλημα "λειτουργεί στον υπολογιστή μου".
- **Απομόνωση:** Κάθε υπηρεσία λειτουργεί στο δικό της περιβάλλον, ελαχιστοποιώντας τις συγκρούσεις μεταξύ των υπηρεσιών και απλοποιώντας τον εντοπισμό σφαλμάτων και τη συντήρηση.
- **Επεκτασιμότητα:** Τα δοχεία Docker μπορούν εύκολα να κλιμακωθούν προς τα πάνω ή προς τα κάτω με βάση το φορτίο, καθιστώντας το σύστημα προσαρμόσιμο σε ποικίλες απαιτήσεις.

- **Φορητότητα:** Τα containers μπορούν να αναπτυχθούν σε οποιοδήποτε σύστημα υποστηρίζει Docker, ενισχύοντας την ευελιξία των επιλογών ανάπτυξης.

Η ανάπτυξη περιλαμβάνει τη δημιουργία εικόνων Docker για κάθε υπηρεσία, οι οποίες στη συνέχεια εκτελούνται ως container στον κεντρικό υπολογιστή ή στην υποδομή cloud. Το Docker Compose μπορεί να χρησιμοποιηθεί για τη διαχείριση ρυθμίσεων πολλών container, καθορίζοντας τον τρόπο με τον οποίο αυτά τα container αλληλεπιδρούν μεταξύ τους και διασφαλίζοντας συντονισμένη εκκίνηση και τερματισμό λειτουργίας.

Η χρήση του Docker για την ανάπτυξη των υπηρεσιών σε αυτό το έργο όχι μόνο βελτιστοποιεί τη διαδικασία ανάπτυξης και ανάπτυξης, αλλά ενισχύει επίσης την αξιοπιστία και την επεκτασιμότητα του συστήματος. Αυτή η προσέγγιση ευθυγραμμίζεται με τις σύγχρονες βέλτιστες πρακτικές για την ανάπτυξη λογισμικού, διασφαλίζοντας ότι το αναπτυχθέν σύστημα είναι σε καλή θέση για μελλοντική ανάπτυξη και βελτιώσεις.

5. Σχεδιασμός Μονάδων Συστήματος

5.1. Μονάδα Συλλογής Δεδομένων (Twitter & Reddit)

Η Μονάδα Συλλογής Δεδομένων (Data Collection Module) είναι ένα κρίσιμο στοιχείο του συστήματος, υπεύθυνο για τη συλλογή σχετικών δεδομένων από δύο μεγάλες πλατφόρμες μέσων κοινωνικής δικτύωσης: το Twitter και το Reddit. Αυτή η ενότητα χρησιμοποιεί τα αντίστοιχα APIs αυτών των πλατφορμών για πρόσβαση και ανάκτηση δεδομένων.

Βασικά χαρακτηριστικά:

- **Ενσωμάτωση API Twitter:**
 - Χρησιμοποιεί το API του Twitter για τη λήψη tweets με βάση συγκεκριμένες λέξεις-κλειδιά, hashtags ή προφίλ χρηστών που σχετίζονται με τα απόβλητα τροφίμων.
 - Προσφέρει ροή δεδομένων σε πραγματικό χρόνο καθώς και πρόσβαση σε ιστορικά tweets.
 - Χειρίζεται τα όρια ρυθμού και τους περιορισμούς API για να διασφαλίσει τη συνεχή συλλογή δεδομένων.
- **Ενσωμάτωση Reddit API:**
 - Αξιοποιεί το API του Reddit για εξαγωγή αναρτήσεων και σχολίων από σχετικά subreddits ή νήματα.
 - Εστιάζει σε θέματα σχετικά με τη σπατάλη τροφίμων, τη βιωσιμότητα και σχετικές συζητήσεις.
 - Εξασφαλίζει αποτελεσματική ανάκτηση δεδομένων με ταυτόχρονη τήρηση των οδηγιών API του Reddit.
- **Χειρισμός δεδομένων:**
 - Και τα δύο API επιστρέφουν δεδομένα σε μορφή JSON, τα οποία στη συνέχεια αναλύονται και δομούνται για συνέπεια.
 - Οι βασικές πληροφορίες που εξάγονται περιλαμβάνουν περιεχόμενο κειμένου, χρονικές σημάνσεις, πληροφορίες χρήστη, μετρήσεις αφοσίωσης (όπως retweet, επισημάνσεις "μου αρέσει" ή θετικές ψήφους) και άλλα σχετικά μεταδεδομένα.
- **Επεκτασιμότητα και Αξιοπιστία:**
 - Χειρίζεται αποτελεσματικά μεγάλους όγκους δεδομένων.
 - Περιλαμβάνει μηχανισμούς χειρισμού σφαλμάτων και ανάκτησης για τη διαχείριση πιθανών χρόνων διακοπής λειτουργίας ή αλλαγών API.
- **Απόρρητο και συμμόρφωση:**
 - Διασφαλίζει τη συμμόρφωση με τους όρους παροχής υπηρεσιών τόσο του Twitter όσο και του Reddit.
 - Σέβεται τις πολιτικές απορρήτου των χρηστών και χρήσης δεδομένων, όπως περιγράφονται από τις πλατφόρμες.
- **Ενσωμάτωση με MongoDB:**

- Τα δεδομένα που συλλέγονται αποθηκεύονται στη MongoDB, επιτρέποντας την ευέλικτη και επεκτάσιμη αποθήκευση.
- Η φύση της MongoDB χωρίς αυστηρό σχήμα δεδομένων φιλοξενεί εύκολα και αποδοτικά τις ποικίλες δομές δεδομένων από το Twitter και το Reddit.

5.2. Μονάδα Προεπεξεργασίας Δεδομένων

5.2.1. Μετάφραση κειμένου στην Αγγλική Γλώσσα

Χρησιμοποιείται για τη μετατροπή όλων των δεδομένων κειμένου στα αγγλικά, διασφαλίζοντας ομοιομορφία στη γλώσσα για συνεπή επεξεργασία και ανάλυση.

- **Ανίχνευση γλώσσας**
 - Πριν από τη μετάφραση, είναι απαραίτητο να προσδιοριστεί η γλώσσα κάθε κειμένου. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας βιβλιοθήκες ανίχνευσης γλωσσών όπως το langdetect στην Python.
 - Κάθε κομμάτι κειμένου περνά από τον αλγόριθμο ανίχνευσης γλώσσας για να προσδιοριστεί η τρέχουσα γλώσσα του.
- **Μετάφραση**
 - Τα κείμενα που προσδιορίζονται ως μη αγγλικά μεταφράζονται στη συνέχεια στα αγγλικά.
 - Αυτό πραγματοποιείται χρησιμοποιώντας το μεταφραστικό API Google Translate.
 - Η διαδικασία μετάφρασης περιλαμβάνει την αποστολή του πρωτότυπου κειμένου στη μεταφραστική υπηρεσία και τη λήψη του μεταφρασμένου αγγλικού κειμένου ως απάντηση.

5.2.2. Μετατροπή κειμένου σε πεζά

- **Μαθηματική Διατύπωση:**
 - Έστω X το σύνολο όλων των χαρακτήρων του κειμένου.
 - Για κάθε χαρακτήρα $x \in X$, μετατρέπεται το x στο πεζό ισοδύναμό του x_{lower} .
- **Λεπτομέρειες:**

Η διαδικασία είναι επαναληπτική για κάθε χαρακτήρα του κειμένου και την εφαρμογή του μετασχηματισμού πεζών. Εξασφαλίζει ομοιομορφία στο κείμενο, καθιστώντας το χωρίς διάκριση πεζών-κεφαλαίων, κάτι που είναι ζωτικής σημασίας για συνεπή ανάλυση κειμένου.

5.2.3. Αφαίρεση Ετικετών HTML

- **Μαθηματική Διατύπωση:**
 - Έστω X το αρχικό κείμενο με ετικέτες HTML.
 - Ορίζεται μια συνάρτηση $f(x)$ που προσδιορίζει και αφαιρεί ετικέτες HTML από το x .
 - Το κείμενο X' που προκύπτει είναι το X μετά την εφαρμογή του $f(x)$.
- **Λεπτομέρειες:**

Οι ετικέτες HTML στο κείμενο είναι μη κειμενικά στοιχεία που μπορούν να παραμορφώσουν την ανάλυση κειμένου. Η συνάρτηση $f(x)$ σαρώνει το κείμενο για μοτίβα ετικετών HTML και τα αφαιρεί, αφήνοντας πίσω μόνο το περιεχόμενο κειμένου.

5.2.4. Επέκταση Συντομογραφιών

– **Μαθηματική Διατύπωση:**

- Έστω X το σύνολο των λέξεων του κειμένου.
- Ορίζεται μια συνάρτηση αντιστοίχισης $f(x) \rightarrow Y$ όπου Y είναι το σύνολο των διευρυμένων μορφών λέξεων.
- Για κάθε λέξη $x \in X$, εάν το x είναι συντομογραφία, αντικαθιστά το x με την $f(x)$.

– **Λεπτομέρειες:**

Οι συσπάσεις είναι συντομευμένες μορφές λέξεων ή φράσεων (όπως «δεν γίνεται» για «δεν»). Η επέκτασή τους βοηθά στη σαφήνεια και την ομοιομορφία, ειδικά για εργασίες επεξεργασίας φυσικής γλώσσας. Η συνάρτηση αντιστοίχισης f χρησιμοποιείται για τη συστηματική επέκταση αυτών των συστολών.

5.2.5. Μετατροπή αριθμών σε λέξεις

– **Μαθηματική Διατύπωση:**

- Έστω X το σύνολο όλων των στοιχείων (λέξεων και αριθμών) του κειμένου.
- Ορίζεται μια συνάρτηση $g(x)$ που μετατρέπει αριθμητικά στοιχεία στο X στα ισοδύναμα λέξης τους.
- Εφαρμόζεται το $g(x)$ σε κάθε στοιχείο στο X για να ληφθεί το X' , όπου όλοι οι αριθμοί είναι σε μορφή λέξης.

– **Λεπτομέρειες:**

Τα αριθμητικά δεδομένα στο κείμενο μπορεί να είναι πιο σημαντικά όταν μετατρέπονται σε λέξεις, ειδικά σε ορισμένα αναλυτικά πλαίσια. Η συνάρτηση $g(x)$ προσδιορίζει αριθμητικά μοτίβα και τα αντικαθιστά με τις αντίστοιχες αναπαραστάσεις λέξεων.

5.2.6. Αφαίρεση ειδικών χαρακτήρων

– **Μαθηματική Διατύπωση:**

- Έστω X το αρχικό κείμενο που περιέχει ειδικούς χαρακτήρες.
- Ορίζεται μια συνάρτηση $h(x)$ που προσδιορίζει και αφαιρεί ειδικούς χαρακτήρες από το x .
- Το κείμενο που προκύπτει X' είναι το X μετά την εφαρμογή του $h(x)$.

– **Λεπτομέρειες:**

- Ειδικοί χαρακτήρες όπως σημεία στίξης, σημεία τονισμού και διακριτικά μπορεί να είναι άσχετοι ή ενοχλητικοί στην ανάλυση κειμένου. Η συνάρτηση $h(x)$ αφαιρεί αυτούς τους χαρακτήρες, απλοποιώντας το κείμενο για περαιτέρω επεξεργασία.

5.2.7. Αφαίρεση λευκών διαστημάτων

– **Μαθηματική Διατύπωση:**

- Έστω X το κείμενο με πιθανά εξωτερικά κενά.
- Ορίζεται μια συνάρτηση $k(x)$ που περικόπτει τα κύρια, υστερούντα και πλεονάζοντα εσωτερικά λευκά κενά.
- Το καθαρισμένο κείμενο X' προκύπτει με την εφαρμογή του $k(x)$ στο X .

– **Λεπτομέρειες:**

Τα υπερβολικά λευκά κενά, είτε στην αρχή, είτε στο τέλος ή εντός του κειμένου, μπορεί να επηρεάσουν την ακρίβεια της ανάλυσης κειμένου. Η συνάρτηση $k(x)$ τυποποιεί τη χρήση του λευκού χώρου στο κείμενο.

5.2.8. Tokenization Word

– **Μαθηματική Διατύπωση:**

- Έστω X το σύνολο των χαρακτήρων που σχηματίζουν το κείμενο.
- Ορίζεται μια συνάρτηση tokenization $j(x)$ που χωρίζει το X σε ένα σύνολο διακριτικών T με βάση προκαθορισμένους οριοθέτες (όπως κενά).

– **Λεπτομέρειες:**

- Tokenization είναι η διαδικασία διάσπασης του κειμένου σε μικρότερες μονάδες (tokens), συνήθως λέξεις. Αυτό είναι ένα θεμελιώδες βήμα στην ανάλυση κειμένου, που επιτρέπει την εξέταση μεμονωμένων λέξεων ή φράσεων.

5.2.9. Θεματοποίηση και Λημματοποίηση

– **Μαθηματική Διατύπωση:**

- Έστω T το σύνολο των κουπονιών που λαμβάνονται από το tokenization.
- Καθορίζονται οι συναρτήσεις $k(x)$ για το stemming και $l(x)$ για τη λημματοποίηση, που εφαρμόζονται σε κάθε διακριτικό $t \in T$.
- Το προκύπτον σύνολο διακριτικών T' περιέχει βασικές ή λημματοποιημένες μορφές των αρχικών διακριτικών.

– **Λεπτομέρειες:**

Η θεματοποίηση μειώνει τις λέξεις στη ρίζα τους, οδηγώντας συχνά σε χονδροειδή τεμαχισμό λέξεων. Η λημματοποίηση, από την άλλη πλευρά, μειώνει τις λέξεις στη βάση τους ή τη μορφή λεξικού. Και οι δύο διαδικασίες βοηθούν στη μείωση της πολυπλοκότητας των δεδομένων κειμένου για ανάλυση.

5.2.10. Αφαίρεση κοινών λέξεων

– **Μαθηματική Διατύπωση:**

- Έστω T' το σύνολο των κουπονιών μετά το stemming/lemmatization.
- Καθορίζεται ένα σύνολο S που περιέχει κοινές λέξεις.
- Το τελικό σύνολο κουπονιών T'' λαμβάνεται αφαιρώντας οποιοδήποτε διακριτικό στο T' που βρίσκεται επίσης στο S .

– **Λεπτομέρειες:**

Απομακρύνονται οι κοινές λέξεις όπως "και", "το", "είναι" κ.λπ., που δεν συνεισφέρουν στη δεικτοδότηση κειμένου. Η κατάργηση αυτών των λέξεων καθιστά πιο αποδοτική και ακριβή την δεικτοδότηση/διαχείριση του κειμένου. Κάθε ένα από αυτά τα βήματα συμβάλλει στον καθαρισμό και την τυποποίηση των δεδομένων κειμένου, καθιστώντας το πιο κατάλληλο για ακριβή ανάλυση, όπως ανάλυση συναισθήματος ή μοντελοποίηση θεμάτων.

5.3. Μονάδα συστήματος βάσης δεδομένων - Ρύθμιση και σχήμα MongoDB

5.3.1. Εγκατάσταση και διαμόρφωση

Στην ενότητα αυτή αναδεικνύονται οι προηγμένες λειτουργικότητες της MongoDB, οι οποίες αποτελούν σημαντικό μέρος του έργου. Αρχικά, η MongoDB έχει εγκατασταθεί και ρυθμιστεί για να λειτουργεί ως υπηρεσία, παρέχοντας συνεχή διαθεσιμότητα για την αποθήκευση και ανάκτηση δεδομένων. Αυτό είναι ουσιώδες για την διασφάλιση της αδιάλειπτης λειτουργίας και πρόσβασης στα δεδομένα, κρίσιμα για την επιτυχία του έργου.

Παράλληλα, η βάση δεδομένων της MongoDB προσφέρει ένα υψηλό επίπεδο ασφάλειας, με την ενσωμάτωση μηχανισμών ελέγχου πρόσβασης και την προστασία της ακεραιότητας και του απορρήτου των δεδομένων. Αυτό είναι ζωτικής σημασίας σε έναν κόσμο όπου η ασφάλεια των δεδομένων αποτελεί συνεχώς αυξανόμενη ανησυχία, εξασφαλίζοντας ότι τα δεδομένα παραμένουν ασφαλή και προστατευμένα.

5.3.2. Σχεδιασμός Σχήματος

Η δομή της Βάσης Δεδομένων που περιγράφεται φαίνεται να είναι εξαιρετικά οργανωμένη και λειτουργική, ιδιαίτερα στοχευμένη για την αποδοτική διαχείριση και ανάλυση δεδομένων από διάφορες πηγές. Η σχεδίαση και δόμηση της βάσης σε διαφορετικές συλλογές επιτρέπει τον λογικό διαχωρισμό των δεδομένων, εξασφαλίζοντας ότι κάθε κατηγορία δεδομένων είναι εύκολα προσβάσιμη και διαχειρίσιμη.

Οι συλλογές όπως «tweets», «reddits» και «processed_results» αντιπροσωπεύουν διαφορετικές πηγές και είδη δεδομένων, κάθε μία με τον δικό της ρόλο στη γραμμή δεδομένων. Για παράδειγμα, τα «tweets» και τα «reddits» περιέχουν αρχικά δεδομένα από τα αντίστοιχα μέσα κοινωνικής δικτύωσης, ενώ τα «processed_results» περιλαμβάνουν δεδομένα που έχουν υποστεί επεξεργασία και ανάλυση. Αυτή η διαχωριστική προσέγγιση διευκολύνει την αποδοτική διαχείριση και ανάλυση των δεδομένων, ενισχύοντας την οργάνωση και την επεξεργασία τους.

Δυναμικό σχήμα

Η απουσία σχήματος επιτρέπει την ευέλικτη αποθήκευση δεδομένων, καλύπτοντας την ποικιλόμορφη και δυναμική φύση των δεδομένων μέσων κοινωνικής δικτύωσης. Τα έγγραφα εντός των συλλογών δεν χρειάζεται να έχουν ομοιόμορφη δομή, επιτρέποντας την αποθήκευση ποικίλων μορφών δεδομένων από το Twitter και το Reddit.

Τα Βασικότερα πεδία στις συλλογές

- Συλλογή `tweets`: Αποθηκεύει ακατέργαστα tweets, συμπεριλαμβανομένων πεδίων όπως "id", "text", "created_at", "user", "hashtags" και "αναφορές".
- Συλλογή `reddits`: Περιέχει αναρτήσεις Reddit με πεδία όπως `id`, `title`, `body`, `created\_at` και `subreddit`.
- Συλλογή `processed\_results`: Περιέχει τα αποτελέσματα της επεξεργασίας, συμπεριλαμβανομένων των βαθμολογιών συναισθημάτων, συχνών λέξεων και αποτελεσμάτων μοντελοποίησης θεμάτων.

Ευρετηρίαση

Η στρατηγική ευρετηρίασης εφαρμόζεται για τη βελτιστοποίηση της απόδοσης του ερωτήματος, ειδικά για πεδία στα οποία γίνεται συχνή πρόσβαση ή ερωτήματα, όπως το "id" και το "created_at".

Ρόλος στην προσωρινή αποθήκευση (caching)

Η MongoDB διαδραματίζει κρίσιμο ρόλο στην προσωρινή αποθήκευση υπολογισμένων αποτελεσμάτων για τη βελτίωση της απόδοσης του πίνακα εργαλείων. Μια ξεχωριστή συλλογή, "cache", χρησιμοποιείται για την αποθήκευση προυπολογισμένων δεδομένων, μειώνοντας το χρόνο φόρτωσης για πολύπλοκες απεικονίσεις.

Επεκτασιμότητα και απόδοση

Η αρχιτεκτονική της MongoDB, όπως περιγράφεται εδώ, παρουσιάζει εξαιρετικές δυνατότητες που την καθιστούν ιδανική για την διαχείριση και ανάλυση μεγάλων όγκων δεδομένων, ειδικά από κοινωνικά μέσα. Η υποστήριξη της επεκτασιμότητας είναι ένα κρίσιμο χαρακτηριστικό, καθώς αυτό επιτρέπει στη βάση δεδομένων να διαχειρίζεται αποτελεσματικά αυξανόμενους όγκους δεδομένων χωρίς απώλεια απόδοσης.

Ο συντονισμός απόδοσης μέσω της προσαρμογής των προτιμήσεων ανάγνωσης/εγγραφής και του διαμοιρασμού διασφαλίζει ότι η βάση δεδομένων διατηρεί τη βέλτιστη απόδοση, ακόμη και κάτω από αυξημένες απαιτήσεις. Αυτό είναι ιδιαίτερα σημαντικό σε εφαρμογές όπου η γρήγορη ανάκτηση και επεξεργασία δεδομένων είναι κρίσιμη. Η τακτική δημιουργία αντιγράφων ασφαλείας και οι στρατηγικές ανάκτησης δεδομένων παρέχουν επιπλέον επίπεδα ασφάλειας, διασφαλίζοντας ότι οι πληροφορίες παραμένουν ασφαλείς και μπορούν να επαναφερθούν σε περίπτωση αστοχίας συστήματος ή καταστροφής δεδομένων.

Τέλος, η ευέλικτη και επεκτάσιμη φύση της MongoDB την καθιστά ιδανική για τον χειρισμό των δυναμικών και ποικίλων δεδομένων από πλατφόρμες κοινωνικών μέσων. Η ικανότητά της να οργανώνει και διαχειρίζεται κείμενα, μαζί με την ταχύτητα ανάκτησης και την αποδοτική διαχείριση μεγάλων συνόλων δεδομένων, την καθιστούν μια εξαιρετική επιλογή για τις ανάγκες του έργου.

5.4. Μονάδα οπτικοποίησης των αποτελεσμάτων με τη χρήση του Dash Plotly

Το Dash Plotly αντιπροσωπεύει μια ισχυρή πλατφόρμα για τη δημιουργία διαδραστικών εφαρμογών ιστού, με έμφαση στην οπτικοποίηση δεδομένων. Ως πλαίσιο ανοιχτού κώδικα, προσφέρει την ευελιξία και την δυνατότητα προσαρμογής που είναι σημαντική για τέτοιου είδους εφαρμογές. Η ενσωμάτωσή του με την Python για τον χειρισμό δεδομένων αποτελεί ένα βασικό χαρακτηριστικό, καθώς η Python είναι μια δημοφιλής γλώσσα προγραμματισμού με ισχυρές βιβλιοθήκες για την ανάλυση και την οπτικοποίηση δεδομένων.

Το γεγονός ότι το dashboard έχει σχεδιαστεί με μια φιλική προς το χρήστη διεπαφή είναι κρίσιμο, καθώς αυτό βοηθά στην εύκολη πλοήγηση και αλληλεπίδραση με τις απεικονίσεις. Η προσεκτική επιλογή των αισθητικών και λειτουργικών στοιχείων συμβάλλει στη βελτίωση της εμπειρίας του χρήστη, ενώ ταυτόχρονα διασφαλίζει την καθαρότητα και την αναγνωσιμότητα των δεδομένων.

Η δυναμική σύνδεση του πίνακα εργαλείων με τη βάση δεδομένων MongoDB είναι επίσης σημαντική, καθώς επιτρέπει την άμεση οπτικοποίηση δεδομένων που λαμβάνονται σε πραγματικό χρόνο. Η χρήση δεδομένων από συλλογές όπως τα «processed_results» και η «κρυφή μνήμη» για τη δημιουργία διαγραμμάτων και γραφημάτων αποδεικνύει την πολυπλοκότητα και την ικανότητα του συστήματος να διαχειρίζεται και να αναλύει πολύπλοκα σύνολα δεδομένων. Αυτό καθιστά το Dash Plotly μια ιδανική λύση για την παρουσίαση και την ανάλυση δεδομένων σε ένα εύχρηστο, διαδραστικό και ενημερωτικό τρόπο.

5.4.1. Βασικές Οπτικοποιήσεις

- **Ανάλυση συναισθήματος:** Γραφήματα που εμφανίζουν τάσεις συναισθήματος με την πάροδο του χρόνου, κατανομή συναισθήματος σε διαφορετικές πλατφόρμες μέσων κοινωνικής δικτύωσης και βαθμολογίες συναισθήματος που σχετίζονται με συγκεκριμένες λέξεις-κλειδιά.
- **Ανάλυση συχνότητας:** Διαγράμματα που δείχνουν τη συχνότητα συγκεκριμένων λέξεων ή φράσεων, τονίζοντας τα διαδεδομένα θέματα στη συζήτηση των μέσων κοινωνικής δικτύωσης σχετικά με τη σπατάλη τροφίμων.
- **Μοντελοποίηση θεμάτων:** Διαδραστικές απεικονίσεις, όπως σύννεφα λέξεων ή γραφήματα δικτύου, που αντιπροσωπεύουν τα αποτελέσματα της μοντελοποίησης θεμάτων, αποκαλύπτοντας κυρίαρχα θέματα και τις διασυνδέσεις τους.

5.4.2. Βασικές Λειτουργίες

Οι χρήστες μπορούν να αλληλεπιδράσουν με τις απεικονίσεις, όπως φιλτράρισμα δεδομένων με βάση το εύρος ημερομηνιών, τις λέξεις-κλειδιά ή τις πλατφόρμες. Οι συμβουλές εργαλείων και τα στοιχεία με δυνατότητα κλικ παρέχουν πρόσθετες πληροφορίες, ενισχύοντας τις αναλυτικές δυνατότητες του πίνακα εργαλείων.

Οι μηχανισμοί προσωρινής αποθήκευσης εφαρμόζονται για την αποθήκευση προϋπολογισμένων αποτελεσμάτων για σύνθετες απεικονίσεις, διασφαλίζοντας ταχύτερους χρόνους φόρτωσης και ομαλότερη εμπειρία χρήστη. Το dashboard είναι βελτιστοποιημένο για διάφορα μεγέθη οθόνης και συσκευές, διασφαλίζοντας προσβασιμότητα και σταθερή

απόδοση. Το dashboard έχει σχεδιαστεί για αυτόματη ανανέωση σε τακτά χρονικά διαστήματα, διασφαλίζοντας την εμφάνιση των πιο πρόσφατων δεδομένων. Παρέχονται επίσης επιλογές μη αυτόματης ανανέωσης για τους χρήστες να ενημερώνουν τα δεδομένα κατά παραγγελία.

Ο πίνακας εργαλείων Dash Plotly χρησιμεύει ως η κεντρική πλατφόρμα για την οπτικοποίηση και την ερμηνεία των αναλυόμενων δεδομένων κοινωνικών μέσων. Παρέχει στους ενδιαφερόμενους φορείς αξιόπιστες γνώσεις σχετικά με την αντίληψη του κοινού και τις τάσεις που σχετίζονται με τη σπατάλη τροφίμων, βοηθώντας στη λήψη αποφάσεων και στη διαμόρφωση στρατηγικής. Η διαδραστική και διαισθητική φύση του dashboard τον καθιστά αποτελεσματικό εργαλείο για την επικοινωνία σύνθετων δεδομένων σε κατανοητή μορφή.

5.5. Σχεδιασμός της στρατηγικής προσωρινής αποθήκευσης (cache DB)

Ο σκοπός της εν λόγω μεθόδου είναι να βελτιώσει την απόδοση και την αποδοτικότητα των στοιχείων επεξεργασίας και οπτικοποίησης δεδομένων στο αναπτυχθέν σύστημα. Δεδομένου του μεγάλου όγκου δεδομένων και του υψηλού φόρτου υπολογισμών που απαιτείται σε κάποιες μονάδες του συστήματος (όπως ο προσδιορισμός θεμάτων και η ανάλυση συναισθήματος), η προσωρινή αποθήκευση γίνεται μια κρίσιμη πτυχή για τη διασφάλιση ενός υψηλά αλληλεπιδραστικού και αποτελεσματικού συστήματος.

Ο μηχανισμός cache υλοποιείται μέσω του κάτωθι βασικού σχεδιασμού:

- **Αποθήκευση προϋπολογισμένων αποτελεσμάτων:**
 - Λειτουργία: Αποθήκευση των αποτελεσμάτων χρονοβόρων υπολογισμών (όπως μοντελοποίηση θεμάτων και ανάλυση co-occurrence) σε ξεχωριστή συλλογή MongoDB.
 - Πλεονέκτημα: Σημαντική μείωση του χρόνου φόρτωσης για τον πίνακα εργαλείων, καθώς το σύστημα ανακτά προϋπολογισμένα αποτελέσματα αντί να τα υπολογίζει εκ νέου για κάθε περίοδο λειτουργίας χρήστη.
- **Προσωρινή αποθήκευση βάσει χρονικής σφραγίδας (timestamp):**
 - Μέθοδος: Κάθε αποθηκευμένο αποτέλεσμα στην κρυφή μνήμη σχετίζεται με μια χρονική σήμανση που υποδεικνύει πότε ενημερώθηκε για τελευταία φορά.
 - Χρήση: Το σύστημα ελέγχει τη χρονική σήμανση για να αποφασίσει εάν θα χρησιμοποιήσει τα αποθηκευμένα δεδομένα ή θα υπολογίσει εκ νέου τα αποτελέσματα, διασφαλίζοντας ότι τα δεδομένα που παρουσιάζονται είναι ενημερωμένα.
- **Επιλεκτική προσωρινή αποθήκευση:**
 - Στρατηγική: Αποθηκεύονται μόνο συγκεκριμένες, υπολογιστικά εντατικές λειτουργίες. Για παράδειγμα, τα αποτελέσματα της ανάλυσης συναισθήματος και της ανάλυσης συχνότητας των λέξεων αποθηκεύονται στην κρυφή μνήμη λόγω της υψηλής υπολογιστικής τους απαίτησης.
 - Αποτέλεσμα: Αυτή η επιλεκτική προσέγγιση βελτιστοποιεί τη χρήση των πόρων, αποθηκεύοντας μόνο ότι είναι απαραίτητο και υπολογίζοντας δυναμικά εκ νέου τα λιγότερο εντατικά δεδομένα.

– **Ακύρωση και ενημέρωση προσωρινής μνήμης :**

- Διαδικασία: Εφαρμογή ενός μηχανισμού για την ακύρωση και την ενημέρωση της προσωρινής μνήμης όταν προστίθενται νέα δεδομένα στη βάση δεδομένων ή όταν επιτυγχάνεται ένα συγκεκριμένο χρονικό όριο.
- Λειτουργία: Διασφαλίζει ότι τα αποθηκευμένα δεδομένα παραμένουν σχετικά και αντικατοπτρίζουν τις πιο πρόσφατες πληροφορίες.
- Η συγκεκριμένα μονάδα συνετέλεσε σε σημαντικό βαθμό στην:
- Βελτίωση απόδοσης : Η προσωρινή αποθήκευση μειώνει σημαντικά τον χρόνο απόκρισης του dashboard, βελτιώνοντας την εμπειρία χρήστη.
- Βελτιστοποίηση πόρων: Με την αποφυγή περιττών υπολογισμών, οι πόροι του συστήματος χρησιμοποιούνται πιο αποτελεσματικά, επιτρέποντας ομαλότερη λειτουργία ακόμη και με μεγάλα σύνολα δεδομένων.
- Επεκτασιμότητα: Η στρατηγική προσωρινής αποθήκευσης συμβάλλει στην επεκτασιμότητα του συστήματος, καθώς μπορεί να χειριστεί αυξημένα φορτία δεδομένων χωρίς ανάλογη αύξηση του χρόνου επεξεργασίας.

5.5.1. Τεχνική Υλοποίηση

- Μηχανισμός ενημέρωσης προσωρινής μνήμης : Εφαρμογή συστήματος όπου η κρυφή μνήμη ενημερώνεται τακτικά ώστε να αντικατοπτρίζει τα πιο πρόσφατα δεδομένα, διασφαλίζοντας ακρίβεια και συνάφεια.
- Επιλεκτική προσωρινή αποθήκευση: Προσδιορισμός και αποθήκευση μόνο εκείνων των λειτουργιών και των συνόλων δεδομένων που επηρεάζουν περισσότερο την απόδοση, βελτιστοποιώντας τη χρήση των πόρων.

Η προσωρινή αποθήκευση διαδραματίζει καθοριστικό ρόλο στη βελτίωση της απόδοσης του dashboard μειώνοντας τους χρόνους φόρτωσης, εξισορροπώντας το υπολογιστικό φορτίο και διασφαλίζοντας μια ομαλή εμπειρία χρήστη με απόκριση. Αυτή η προσέγγιση είναι ζωτικής σημασίας για τον αποτελεσματικό και αποτελεσματικό χειρισμό μεγάλων όγκων δεδομένων σε ένα περιβάλλον μεγάλων δεδομένων.

6. Αναλύσεις

6.1. Ανάλυση Συναισθήματος με χρήση του VADER

Το VADER (Valence Aware Dictionary and SEntiment Reasoner) είναι να εργαλείο ανάλυσης συναισθήματος βασισμένο σε λεξικό και κανόνες, ειδικά προσαρμοσμένο στα συναισθήματα που εκφράζονται στα μέσα κοινωνικής δικτύωσης.

Βασικά χαρακτηριστικά:

- Είναι βασισμένο σε Λεξικό: Χρησιμοποιεί ένα λεξικό λέξεων γεμάτο συναισθήματα, κάθε μία από τις οποίες βαθμολογείται για το θετικό ή το αρνητικό της συναίσθημα.
- Είναι βασισμένο σε κανόνες: Εφαρμόζει γραμματικούς και συντακτικούς κανόνες για την κατανόηση των αποχρώσεων του κειμένου, συμπεριλαμβανομένων των κεφαλαίων, των σημείων στίξης και των τροποποιητικών λέξεων.

6.1.1. Προεπεξεργασία

Τα δεδομένα από τα μέσα κοινωνικής δικτύωσης υποβάλλονται πρώτα στα τυπικά βήματα προεπεξεργασίας, όπως μετατροπή σε πεζά, αφαίρεση ετικετών HTML και επέκταση συστολών.

Σημείωση: Το VADER έχει σχεδιαστεί για να λειτουργεί με επάρκεια με το συνολικό κείμενο των μέσων κοινωνικής δικτύωσης. Δεν είναι απαραίτητες οι φάσεις απομάκρυνσης κοινών λέξεων, θεματοποίησης, λημματοποίησης οι οποίες μπορούν ακόμη και να μειώσουν την αποτελεσματικότητα της ανάλυσης συναισθημάτων.

6.1.2. Βαθμολογία Συναισθημάτων

- Κάθε προεπεξεργασμένο κείμενο τροφοδοτείται στο VADER.
- Το VADER αναλύει το κείμενο και ορίζει μια σύνθετη βαθμολογία που συγκεντρώνει το σωρευτικό συναίσθημα του κειμένου.
- **Μαθηματική Διατύπωση:**
 - Η σύνθετη βαθμολογία S_c είναι μια κανονικοποιημένη, σταθμισμένη σύνθετη βαθμολογία που υπολογίζεται αθροίζοντας τις βαθμολογίες σθένους κάθε λέξης στο λεξικό, προσαρμόζεται σύμφωνα με τους κανόνες και στη συνέχεια κανονικοποιείται ώστε να κυμαίνεται μεταξύ -1 (το πιο ακραίο αρνητικό) και +1 (το πιο ακραίο θετικό).
 - Ο τύπος για S_c περιλαμβάνει ένα άθροισμα των βαθμολογιών σθένους μεμονωμένων λέξεων $V(w_i)$, προσαρμοσμένο με ενισχυτές και αρνήσεις και κανονικοποιημένο: $S_c = \text{normalize} \sum_i V(w_i) \times \text{Modifiers}$

6.1.3. Ερμηνεία

Η σύνθετη βαθμολογία μπορεί να κατηγοριοποιηθεί σε διαφορετικούς χαρακτηρισμούς συναισθημάτων (π.χ. θετικός, ουδέτερος, αρνητικός) βάσει προκαθορισμένων ορίων. Αυτή η βαθμολογία χρησιμοποιείται για την κατανόηση της συνολικής αίσθησης μεμονωμένων

κειμένων ή συγκεντρώνεται για την κατανόηση των τάσεων συναισθήματος σε ένα μεγαλύτερο σύνολο δεδομένων.

6.1.4. Πλεονεκτήματα του VADER στην ανάλυση των μέσων κοινωνικής δικτύωσης

- **Ευαισθησία στο θέμα:** Αποτυπώνει αποτελεσματικά το συναίσθημα που εξαρτάται από το πλαίσιο που εκφράζεται μέσω επιλογής λέξεων, σημείων στίξης και κεφαλαίων, που είναι κοινά στα κείμενα των μέσων κοινωνικής δικτύωσης.
- **Χειρισμός Emoticons και Slang:** Αναγνωρίζει και ερμηνεύει αποτελεσματικά τα emoticons, την αργκό και άλλα λεξικά που σχετίζονται με τα μέσα κοινωνικής δικτύωσης.

6.1.5. Εφαρμογή στο Έργο

Στο πλαίσιο της ανάλυσης δεδομένων μέσων κοινωνικής δικτύωσης που σχετίζονται με τη σπατάλη τροφίμων, το VADER υποστηρίζει ικανοποιητικά την κατανόηση του κοινού αισθήματος σχετικά με διάφορες πτυχές της σπατάλης τροφίμων, πρακτικών βιωσιμότητας και συναφών θεμάτων. Η αποτύπωση του συναισθήματος είναι πολύτιμη για τη μέτρηση της ευαισθητοποίησης και των στάσεων του κοινού, είναι δυνατόν να υποστηρίξει αποτελεσματικά τη χάραξη πολιτικής, τις εκστρατείες ευαισθητοποίησης και την περαιτέρω έρευνα στον τομέα.

6.2. Ανάλυση Συχνότητας

Η ανάλυση συχνότητας είναι μια μέθοδος ποσοτικοποίησης της εμφάνισης στοιχείων (όπως λέξεις ή φράσεις) σε ένα σύνολο δεδομένων. Σκοπός της είναι να προσδιορίσει τις πιο συχνά χρησιμοποιούμενες λέξεις ή φράσεις, παρέχοντας πληροφορίες για τα κύρια θέματα (topics) στα δεδομένα κειμένου.

6.2.1. Μεθοδολογία Ανάλυσης Συχνότητας

- **Προετοιμασία δεδομένων:**
 - Μετά τα βήματα προεπεξεργασίας (πεζά γράμματα, αφαίρεση ειδικών χαρακτήρων κ.λπ.), τα δεδομένα κειμένου μετατρέπονται σε μεμονωμένες λέξεις ή φράσεις.
 - Κάθε διακριτικό αντιπροσωπεύει έναν πιθανό υποψήφιο για ανάλυση συχνότητας.
- **Συχνότητες καταμέτρησης:**
 - Δημιουργείται μια κατανομή συχνότητας όπου κάθε μοναδικό διακριτικό μετράται για την εμφάνισή του στο σύνολο δεδομένων.
- **Μαθηματική Διατύπωση:**
 - Έστω $T = \{t_1, t_2, \dots, t_n\}$ το σύνολο των διακριτικών στο σύνολο δεδομένων.
 - Η συχνότητα $F(t)$ ενός διακριτικού t είναι ο αριθμός εμφανίσεων του t στο σύνολο δεδομένων.
 - $F(t) = \text{count}(t) | t \in T$

6.2.2. Ανάλυση και Ερμηνεία

Στη συνέχεια, τα tokens κατατάσσονται με βάση τον αριθμό συχνότητων τους. Τα διακριτικά υψηλής συχνότητας αναλύονται για την κατανόηση κοινών θεμάτων ή θεμάτων στα δεδομένα. Αυτή η ανάλυση μπορεί να οπτικοποιηθεί χρησιμοποιώντας ραβδογράμματα, σύννεφα λέξεων ή άλλες γραφικές αναπαραστάσεις.

6.2.3. Εφαρμογή στο Έργο

Στο πλαίσιο των δεδομένων των μέσων κοινωνικής δικτύωσης για τη σπατάλη τροφίμων, η ανάλυση συχνότητας βοηθά στον εντοπισμό των πιο συζητημένων πτυχών της σπατάλης τροφίμων, όπως συγκεκριμένες πρακτικές, πολιτικές ή προκλήσεις. Παρέχει ένα ποσοτικό μέτρο της προβολής ορισμένων θεμάτων, το οποίο μπορεί να είναι κρίσιμο για την κατανόηση των περιοχών δημόσιου ενδιαφέροντος ή των κενών στην ευαισθητοποίηση.

Για παράδειγμα, μια υψηλή συχνότητα όρων όπως "κομποστοποίηση" ή "ανακύκλωση" μπορεί να υποδηλώνει δημόσια εστίαση σε βιώσιμες πρακτικές, ενώ οι συχνές αναφορές σε "δωρεά τροφής" ή "ανακούφιση από την πείνα" θα μπορούσαν να υπογραμμίσουν ανησυχίες σχετικά με τη διανομή και την ασφάλεια των τροφίμων.

Η συγκεκριμένη τεχνική είναι ιδιαίτερα επεκτάσιμη γεγονός που την καθιστά το κατάλληλο εργαλείο για εφαρμογές μεγάλων δεδομένων. Προσφέρει έναν απλό αλλά ισχυρό τρόπο αποτύπωσης θεμάτων και εξαγωγής ουσιαστικών πληροφοριών από μη δομημένα δεδομένα κειμένου, τα οποία είναι άφθονα στις πλατφόρμες κοινωνικών μέσων.

6.3. Μοντελοποίηση θεμάτων και λανθάνουσα κατανομή Dirichlet (LDA)

Η Μοντελοποίηση θεμάτων είναι μία στατιστική μέθοδος για την ανακάλυψη "κρυμμένων" θεμάτων μέσα σε μια συλλογή εγγράφων. Ο σκοπός της μεθόδου είναι να αποκαλύψει κρυφές θεματικές δομές σε μεγάλα σώματα κειμένου, διευκολύνοντας την κατανόηση μεγάλου όγκου μη δομημένων δεδομένων.

6.3.1. Περιγραφή της Μεθόδου

Η LDA είναι ένα παραγωγικό πιθανοτικό μοντέλο που υποθέτει ότι κάθε έγγραφο είναι ένα μείγμα μικρού αριθμού θεμάτων και ότι κάθε λέξη στο έγγραφο μπορεί να αποδοθεί σε ένα από τα θέματα του εγγράφου. Στην μαθηματική της βάση, το LDA είναι ένα ιεραρχικό Μπεϋζιανό μοντέλο τριών επιπέδων, στο οποίο κάθε αντικείμενο μίας συλλογής μοντελοποιείται ως πεπερασμένο μείγμα από ένα σύνολο θεμάτων. Με τη σειρά του, κάθε θέμα μοντελοποιείται ως ένα άπειρο μείγμα από ένα σύνολο πιθανοτήτων επί των θεμάτων.

– Μηχανισμός Μεθόδου:

- Αρχικοποίηση: Σε κάθε έγγραφο εκχωρείται αρχικά μια τυχαία κατανομή θεμάτων.
- Επαναληπτική διαδικασία:
 - Για κάθε λέξη σε ένα έγγραφο, το LDA προσαρμόζει τις πιθανότητες των θεμάτων σε αυτό το έγγραφο και τις πιθανότητες των λέξεων σε

αυτά τα θέματα, με βάση τις υπόλοιπες λέξεις στο έγγραφο και άλλα έγγραφα.

- Αυτό επιτυγχάνεται μέσω μιας διαδικασίας που ονομάζεται "Δειγματοληψία Gibbs" ή "Variational Inference".

– **Μαθηματική Διατύπωση:**

- Έστω D το σύνολο των εγγράφων και T το σύνολο των θεμάτων.
- Κάθε έγγραφο $d \in D$ αντιπροσωπεύεται ως μια πιθανολογική κατανομή σε θέματα θ_D , όπου $\theta_{t,d}$ είναι η πιθανότητα του θέματος t στο έγγραφο d .
- Κάθε θέμα $t \in T$ αναπαρίσταται ως πιθανολογική κατανομή στις λέξεις φ_t , όπου $\varphi_{t,w}$ είναι η πιθανότητα της λέξης w στο θέμα t .

Ο στόχος είναι να εκτιμηθούν τα θ και φ που μεγιστοποιούν την πιθανότητα των παρατηρούμενων λέξεων στα έγγραφα.

6.3.2. Σημασία στο Πλαίσιο των Μεγάλων Δεδομένων

- **Πολυπλοκότητα χειρισμού:** Η LDA είναι ιδιαίτερα χρήσιμη για την αντιμετώπιση της πολυπλοκότητας και του όγκου των μεγάλων δεδομένων, καθώς μπορεί να επεξεργάζεται μεγάλα σώματα κειμένου αποτελεσματικά.
- **Ακριβέστερες πληροφορίες:** Αποκαλύπτοντας τα λανθάνοντα θέματα, το LDA παρέχει ακριβέστερες πληροφορίες για τα δεδομένα, οι οποίες μπορεί να μην είναι εμφανείς μέσω απλής ανάλυσης λέξεων-κλειδιών ή μετρήσεων συχνότητας.
- **Δυναμική ανάλυση:** Η ικανότητα του LDA να χειρίζεται εξελισσόμενα θέματα με την πάροδο του χρόνου το καθιστά κατάλληλο για την ανάλυση των τάσεων και των αλλαγών της κοινής γνώμης σε πλατφόρμες όπως το Twitter και το Reddit.

6.3.3. Εφαρμογή στο Έργο

Στο πλαίσιο των δεδομένων μέσων κοινωνικής δικτύωσης σχετικά με τα απορρίμματα τροφίμων, η LDA μπορεί να εντοπίσει επικρατέστερα θέματα που συζητούνται σε tweets και αναρτήσεις στο Reddit. Με τη χρήση των Insight, βοηθά στην κατηγοριοποίηση των δεδομένων σε διακριτά θέματα όπως πρακτικές βιωσιμότητας, συζητήσεις πολιτικής, προσωπικές εμπειρίες κ.λπ., παρέχοντας μια δομημένη επισκόπηση του δημόσιου λόγου για τα απόβλητα τροφίμων.

7. Επισκόπηση Δυναμικού Πίνακα Ελέγχου

Ο σκοπός αυτού του πίνακα εργαλείων δεν παρέχει μια ολιστική άποψη του λόγου γύρω από τη σπατάλη τροφίμων και τη βιωσιμότητα στο Twitter. Χρησιμεύει ως ισχυρό εργαλείο για τον εντοπισμό των τάσεων του συναισθήματος, της δημοτικότητας των σχετικών hashtag , της συχνότητας ορισμένων όρων και των συνολικών επιπέδων αφοσίωσης. Συγκεντρώνοντας και οπτικοποιώντας αυτές τις μετρήσεις, ο πίνακας εργαλείων υποστηρίζει τη διαμόρφωση στρατηγικής για εκστρατείες, πρωτοβουλίες πολιτικών και προγραμμάτων ενημέρωσης του κοινού με στόχο τη σπατάλη τροφίμων και την προώθηση βιώσιμων πρακτικών.

7.1. Συνιστώσες

Ο πίνακας εργαλείων αποτελείται από πολλά διαδραστικά στοιχεία, το καθένα σχεδιασμένο για να προσφέρει συγκεκριμένες πληροφορίες:

- **Μετρήσεις Tweet:** Εμφανίζει τον αριθμό των tweet που περιέχουν λέξεις-κλειδιά που σχετίζονται με τα απόβλητα τροφίμων, επιτρέποντας στους χρήστες να μετρούν τον όγκο της συνομιλίας με την πάροδο του χρόνου.
- **Μοναδικές λέξεις και hashtags:** Οπτικοποιεί τον αριθμό των μοναδικών λέξεων και των δημοφιλών hashtag , προσφέροντας πληροφορίες για τους βασικούς όρους και τα θέματα της συζήτησης.
- **Κατανομή συναισθήματος:** Μέσω της ανάλυσης συναισθημάτων, το dashboard αποκαλύπτει τον συναισθηματικό τόνο της συνομιλίας, που κατηγορείται σε θετικά, ουδέτερα και αρνητικά συναισθήματα.
- **Ανάλυση χρονοσειρών:** Δείχνει πώς το συναίσθημα και η αφοσίωση έχουν τάσεις με την πάροδο του χρόνου, αποκαλύπτοντας αιχμές και μειώσεις στην ένταση του λόγου.
- **Συχνότητα λέξεων:** Προσδιορίζει τους όρους που αναφέρονται πιο συχνά στα tweets , ρίχνοντας φως στα κεντρικά σημεία της συνομιλίας.
- **Συνεμφάνιση Hashtag:** Εξετάζει τις σχέσεις μεταξύ διαφορετικών hashtag , απεικονίζοντας τον τρόπο διασύνδεσης διάφορων θεμάτων και καμπανιών.

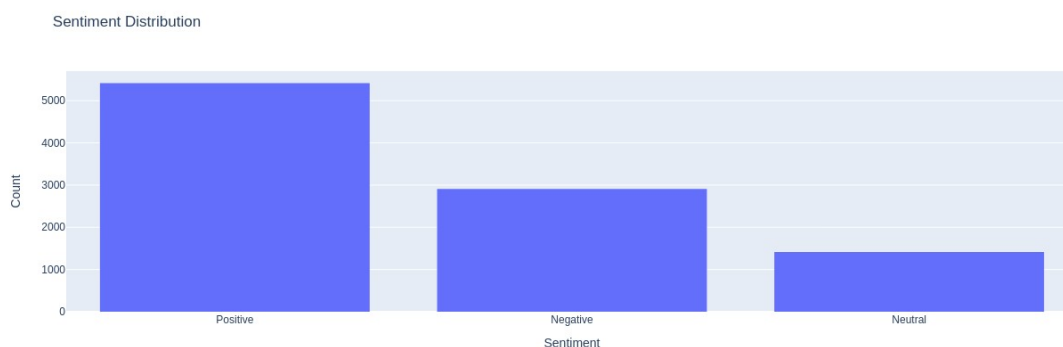
7.2. Επεξεργασία Δεδομένων

Τα υποκείμενα δεδομένα προέρχονται από μια συλλογή MongoDB, όπου κάθε εγγραφή αποτελείται από κείμενο ενός tweet, μεταδεδομένα και προκύπτοντα αναλυτικά στοιχεία, όπως βαθμολογίες συναισθήματος και συχνότητες λέξεων. Τα δεδομένα υποβάλλονται σε επεξεργασία μέσω μιας σειράς συναρτήσεων Python , διασφαλίζοντας ότι μόνο σχετικό, καθαρό κείμενο περιγράφεται στην ανάλυση.

Ο πίνακας εργαλείων χρησιμοποιεί το πλαίσιο Dash του Plotly για τα στοιχεία οπτικοποίησης του front - end, που επιλέγονται για τη διάδραση και την ευκολία χρήσης του. Κάθε γράφημα μπορεί να χειριστεί από τον χρήστη, παρέχοντας μια δυναμική εμπειρία που επιτρέπει τη βαθύτερη εξερεύνηση των δεδομένων. Σχεδιασμένο με γνώμονα την καλή προσβασιμότητα, το

dashboard βασίζεται στον ιστό, διασφαλίζοντας την προσπέλαση από διάφορες συσκευές και πλατφόρμες χωρίς την ανάγκη εξειδικευμένου λογισμικού.

7.3. Ανάλυση Συναισθήματος



Σχήμα 7-1: Κατανομή Συναισθήματος

Περιγραφή

Το γράφημα 1 στον πίνακα εργαλείων, παρέχει μια οπτική αναπαράσταση της κατανομής του συναισθήματος σε όλα τα tweets που σχετίζονται με τη σπατάλη τροφίμων και τη βιωσιμότητα. Χρησιμοποιεί το εργαλείο ανάλυσης συναισθήματος VADER για να ταξινομήσει τα tweets σε τρεις κατηγορίες: θετικά, ουδέτερα και αρνητικά. Το συναίσθημα για κάθε tweet καθορίζεται από το σύνθετο σκορ που δημιουργείται από το VADER, το οποίο είναι μια αθροιστική βαθμολογία θετικότητας, αρνητικότητας και ουδετερότητας.

Λεπτομέρειες Οπτικοποίησης

Η κατανομή συναισθήματος απεικονίζεται μέσω ενός ραβδογράμματος, με τον άξονα x να αντιπροσωπεύει τις κατηγορίες συναισθήματος και τον άξονα y να υποδεικνύει τον αριθμό των tweets. Το ύψος κάθε ράβδου αντιστοιχεί στον αριθμό των tweets που εμπίπτουν στην αντίστοιχη κατηγορία συναισθημάτων. Οι ράβδοι έχουν χρωματική κωδικοποίηση για ευκολία στην ερμηνεία, με το θετικό συναίσθημα να εμφανίζεται συνήθως με πράσινο, το ουδέτερο με γκρι και το αρνητικό με κόκκινο.

Ερμηνεία

Αυτό το γράφημα παρέχει άμεσες πληροφορίες για τον συνολικό συναισθηματικό τόνο της συνομιλίας στο Twitter. Ένας μεγαλύτερος αριθμός θετικών tweets μπορεί να υποδηλώνει ότι η συζήτηση γύρω από τη σπατάλη τροφίμων είναι ελπιδοφόρα και στραμμένη προς το μέλλον, δίνοντας έμφαση σε λύσεις και ιστορίες επιτυχίας. Αντίθετα, η επικράτηση αρνητικού συναισθήματος μπορεί να υποδηλώνει ανησυχίες, κριτικές ή προκλήσεις που κυριαρχούν στη συζήτηση. Ένας μεγάλος αριθμός ουδέτερων tweets θα μπορούσε να υποδηλώνει σημαντική διάδοση πληροφοριών ή ειδήσεων όπου δεν τηρείται ιδιαίτερη συναισθηματική στάση.

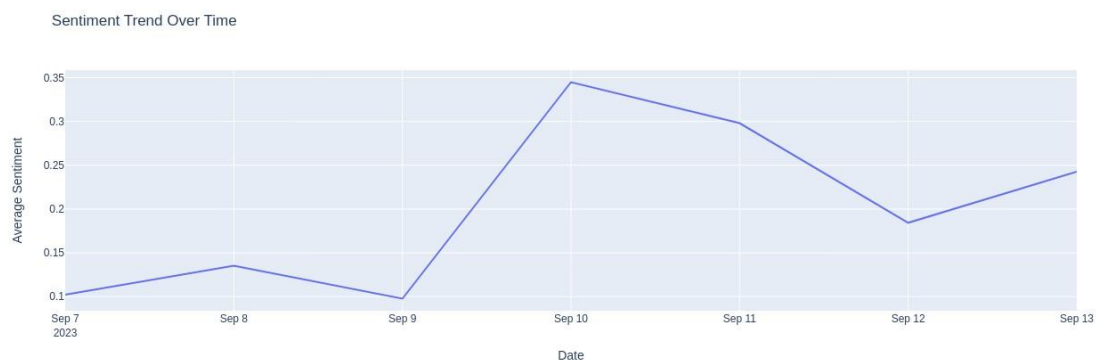
Χρησιμοποίηση

Η κατανόηση της κατανομής συναισθημάτων είναι ζωτικής σημασίας για οργανισμούς και άτομα που στοχεύουν να επηρεάσουν την κοινή γνώμη ή να μετρήσουν το αντίκτυπο των

πρωτοβουλιών τους. Για παράδειγμα, εάν μια πρόσφατη εκστρατεία που στοχεύει στη μείωση της σπατάλης τροφίμων ακολουθείται από την άνοδο του θετικού συναισθήματος, αυτό θα μπορούσε να σηματοδοτήσει αποτελεσματικά μηνύματα. Εναλλακτικά, μια αύξηση του αρνητικού συναισθήματος μπορεί να ωθήσει τους διοργανωτές της καμπάνιας να προσαρμόσουν τη στρατηγική τους.

Παρακολουθώντας τις αλλαγές στο συναίσθημα με την πάροδο του χρόνου, οι συμμετέχοντες μπορούν επίσης να μετρήσουν την αποτελεσματικότητα των αλλαγών πολιτικών, των εκστρατειών ευεξίας του κοινού ή σημαντικών γεγονότων που σχετίζονται με τη βιωσιμότητα των τροφίμων.

7.4. Συναίσθημα με την πάροδο του χρόνου



Σχήμα 7-2: Συναίσθημα με την πάροδο του χρόνου

Περιγραφή

Το γράφημα της Ανάλυσης Συναισθήματος Χρονικής Σειράς παρέχει μια δυναμική απεικόνιση του τρόπου με τον οποίο το δημόσιο αίσθημα για τα θέματα της σπατάλης τροφίμων και της βιωσιμότητας κυμαίνεται με την πάροδο του χρόνου. Η ανάλυση αξιοποιεί ένα γραμμικό διάγραμμα για την παρακολούθηση των βαθμολογιών του σύνθετου συναισθήματος που συγκεντρώνεται σε καθημερινή βάση, εβδομαδιαία ή μηνιαία βάση — ανάλογα με τον όγκο των δεδομένων και την επιθυμητή ευαισθησία των πληροφοριών.

Λεπτομέρειες Οπτικοποίησης

Σε αυτό το γράφημα, ο άξονας x αντιπροσωπεύει το χρόνο, ενώ ο άξονας y ποσοτικοποιεί τη μέση βαθμολογία συναισθήματος για κάθε διάστημα. Κάθε σημείο στο γράφημα γραμμής αντικατοπτρίζει τη μέση βαθμολογία συναισθήματος όλων των tweets που δημοσιεύτηκαν εντός της καθορισμένης περιόδου. Η κίνηση της γραμμής προς τα πάνω ή προς τα κάτω δείχνει μετατοπίσεις προς πιο θετικό ή αρνητικό συναίσθημα, αντίστοιχα.

Ερμηνεία

Οι βασικές γνώσεις μπορούν να εξακολουθήσουν να παρατηρούν τους συσχετισμούς μεταξύ των γεγονότων του πραγματικού κόσμου και αλλαγών στο συναίσθημα. Για παράδειγμα, μια σημαντική αιχμή στο θετικό συναίσθημα μπορεί να ευθυγραμμιστεί με μια μηχανή μείωσης

της πρωτοβουλίας των απορριμμάτων τροφίμων, ενώ ένα κατώτατο σημείο μπορεί να συμπληρωθεί με αναφορές για αυξημένη σπατάλη τροφίμων κατά τις περιόδους διακοπών.

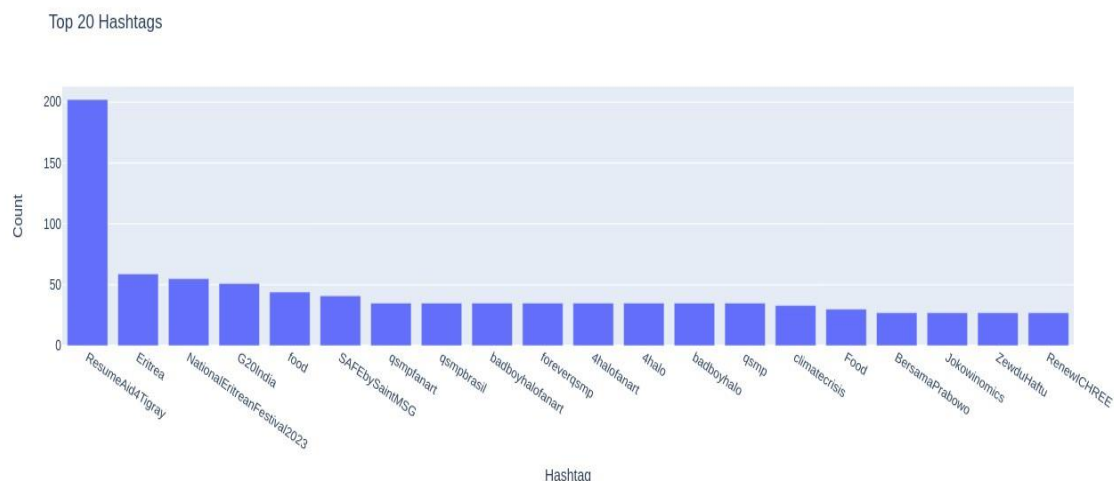
Οι αξιοσημείωτες τάσεις που πρέπει να επισημανθούν θα μπορούσαν να περιλαμβάνουν:

- **Συνεπής θετικότητα:** Μια διαρκής περίοδος θετικού συναισθήματος θα μπορούσε να συνεπάγεται αποτελεσματικές συνεχείς εκστρατείες ή μια αυξανόμενη συλλογική συνείδηση σχετικά με τη βιωσιμότητα.
- **Ξαφνικές αιχμές:** Οι απότομες αυξήσεις στο θετικό ή αρνητικό συναίσθημα μπορούν να συνδέονται με συγκεκριμένα γεγονότα, ανακοινώσεις πολιτικής ή υγειονομικές καμπάνιες στα μέσα κοινωνικής δικτύωσης.
- **Μακροπρόθεσμες τάσεις:** Μια σταδιακή αλλά συνεπής αλλαγή του συναισθήματος για αρκετούς μήνες μπορεί να αντανakλά την αλλαγή της στάσης του κοινού λόγω της εκπαίδευσης, της νομοθεσίας ή της αλλαγής των οικονομικών συνθηκών.

Χρησιμοποίηση

Αυτό το γράφημα χρησιμοποιεί ως στρατηγικό εργαλείο για τους υπεύθυνους χάραξης πολιτικής, τους επαγγελματίες του κοινωνικού μάρκετινγκ και τις ομάδες υπεράσπισης. Ευθυγραμμίζοντας τις τάσεις του συναισθήματος με τα χρονικά διαγράμματα των πρωτοβουλιών τους, οι συμμετέχοντες μπορούν να συμπεράνουν τον αντίκτυπο των ενεργειών τους και να βελτιστοποιήσουν τις μελλοντικές δεσμεύσεις τους. Επιπλέον, ο εντοπισμός τακτικών προτύπων συναισθημάτων μπορεί να βοηθήσει στην πρόβλεψη της δημόσιας απόκρισης σε επαναλαμβανόμενα συμβάντα ή εποχιακά θέματα .

7.5. Ανάλυση Hashtags και Λέξεις



Σχήμα 7-3: Κατάταξη των hashtags

Περιγραφή

Το γράφημα Ανάλυση Hashtags προσφέρει μια οπτική περίληψη των πιο διαδεδομένων hashtags σε συζητήσεις με τη σπατάλη τροφίμων, τη βιωσιμότητα και σχετικά θέματα.

Προβάλλοντας ένα ραβδόγραμμα με μετρήσεις συχνότητας, αυτό το γράφημα επισημαίνει τα hashtags που κυριαρχούν στη συζήτηση μέσα στο σύνολο δεδομένων.

Λεπτομέρειες Οπτικοποίησης

Στο γράφημα, ο άξονας x παραθέτει τα hashtags, ενώ ο άξονας y δείχνει τη συχνότητα εμφάνισης κάθε hashtag. Το μήκος κάθε γραμμής αντιπροσωπεύει τον αριθμό των tweet που περιλαμβάνει το αντίστοιχο hashtag, προσφέροντας μια απλή σύγκριση της σχετικής δημοτικότητάς τους.

Ερμηνεία

Η ανάλυση των κορυφαίων hashtags μπορεί να αποκαλύψει τα κεντρικά σημεία του δημόσιου ενδιαφέροντος και την εμβέλεια της εκστρατείας. Για παράδειγμα:

- **Hashtags καμπάνιας** : Ορισμένα hashtags μπορούν να αντιστοιχούν σε συγκεκριμένες καμπάνιες ή κινήσεις, όπως το #ZeroWaste ή το #FoodRecycle. Η συχνότητά τους μπορεί να υποδεικνύει την εμβέλεια και το επίπεδο δέσμευσης αυτών των πρωτοβουλιών.
- **Hashtags που σχετίζονται με εκδηλώσεις** : Τα hashtags που αυξάνονται στη χρήση μπορούν να σχετίζονται με εκδηλώσεις όπως το #WorldFoodDay ή μια σύνοδος κορυφής για τα απορρίμματα τροφίμων, υποδηλώνοντας τον αντίκτυπο αυτών των γεωγραφικών στη δημόσια συζήτηση.
- **Κοινά θέματα** : Τα επαναλαμβανόμενα θέματα μπορούν να διακριθούν από γενικά αλλά συχνά εμφανιζόμενα hashtags όπως #Sustainability ή #EcoFriendly, υποδεικνύοντας μια γενική ευαισθητοποίηση και ανησυχία για περιβαλλοντικά ζητήματα.
- **Αναδυόμενα θέματα** : Τα νέα ανερχόμενα hashtags θα μπορούσαν να σηματοδοτήσουν αλλαγή εστίασης ή αναδυόμενες τάσεις στη συνομιλία βιωσιμότητας, καθοδηγώντας ενδεχομένως μελλοντικό περιεχόμενο και κατευθύνσεις πολιτικής.

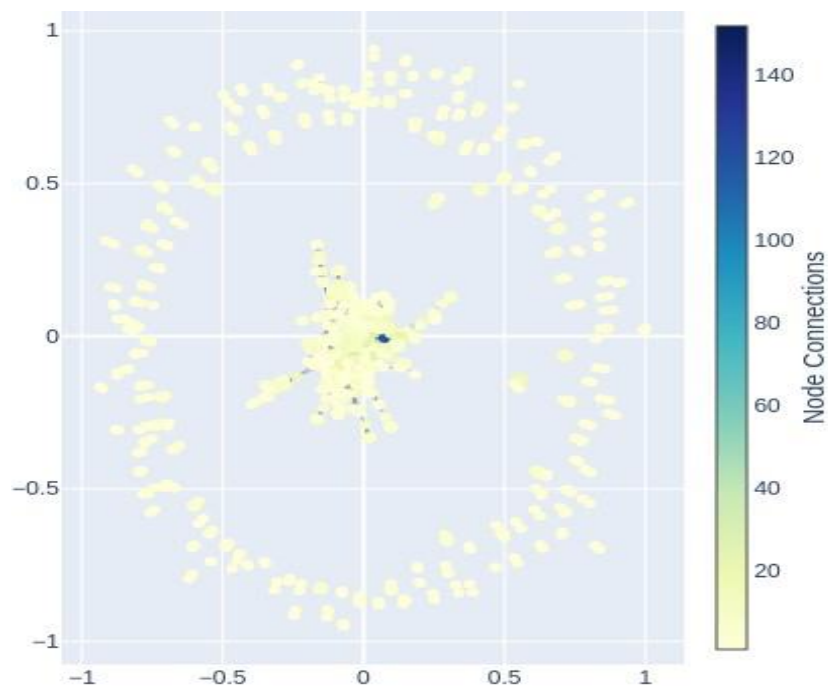
Χρησιμοποίηση

Για τους αναλυτές στρατηγικές μέσων κοινωνικής δικτύωσης και τους περιβαλλοντικούς ακτιβιστές, αυτό το γράφημα λειτουργεί ως μια “σφυγμομέτρηση” σχετικά με τις ποιες συνομιλίες είναι πιο ενδιαφέρουσες ή ευρέως διαδεδομένες. Μπορεί να καθοδηγήσει τη δημιουργία περιεχομένου και τις προσπάθειες υπεράσπισης προς τα θέματα που έχουν μεγαλύτερη απήχηση στο κοινό.

7.6. Δίκτυο συνεμφάνισης Hashtags

Περιγραφή

Το γράφημα του Δικτύου Συνεμφάνισης είναι μια συναρπαστική απεικόνιση που χαρτογραφεί τη διασύνδεση μεταξύ των hashtags που δημοσιεύονται σε tweets σχετικά με τα απόβλητα τροφίμων και σχετικά με θέματα βιωσιμότητας. Αυτό το γράφημα δικτύου απεικονίζει τον τρόπο με τον οποίο ζευγαρώνονται διαφορετικά hashtags μέσα στα tweets, παρέχοντας πληροφορίες για τα συσχετιστικά μοτίβα και τις συμφραζόμενες σχέσεις μεταξύ των εννοιών που αντιπροσωπεύονται από τα hashtag.



Σχήμα 7-4: Δίκτυο Συνεμφάνισης των hastags

Λεπτομέρειες Οπτικοποίησης

Σε αυτό το γράφημα, κάθε κόμβος αντιπροσωπεύει ένα μοναδικό hashtag και οι άκρες (γραμμές) που συνδέουν τους κόμβους αντικατοπτρίζουν τη συνύπαρξη αυτών των hashtags στα ίδια tweets. Το πάχος των άκρων μπορεί να υποδηλώνει τη συχνότητα της ταυτόχρονης εμφάνισης, με παχύτερες γραμμές να υποδηλώνουν μια ισχυρότερη συσχετιστική σχέση. Η τοποθέτηση των κόμβων και η γειτνίασή τους μεταξύ τους μπορεί να προτείνει συστάδες στενά συνδεδεμένων θεμάτων.

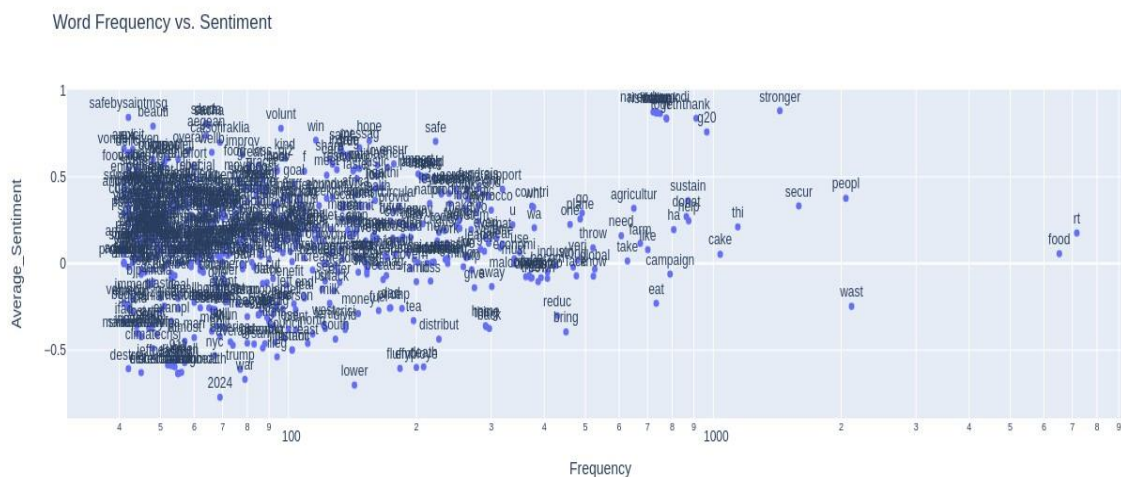
Ερμηνεία

- **Θεματικά συμπλέγματα:** Το γράφημα δικτύου μπορεί να αποκαλύψει συστάδες hashtags που συχνά εμφανίζεται μαζί, υποδεικνύοντας γενικά θέματα στη συζήτηση. Για παράδειγμα, hashtags όπως #FoodRescue, #FoodDonation και #ZeroHunger μπορούν να σχηματίσουν ένα σύμπλεγμα που υποδεικνύει μια συζήτηση γύρω από λύσεις για τη σπατάλη τροφίμων που αντιμετωπίζουν την πείνα.
- **Hashtags με επιρροή:** Οι κεντρικοί κόμβοι με τις περισσότερες συνδέσεις μπορούν να υποδεικνύουν hashtags που είναι καθοριστικής σημασίας στη συζήτηση, χρησιμοποιώντας ως γέφυρες μεταξύ διαφορετικών θεμάτων ή κινήσεων.
- **Εξειδικευμένες συνομιλίες:** Οι περιφερειακοί κόμβοι με λιγότερες συνδέσεις ενδέχεται να προσωποποιήσουν περισσότερα εξειδικευμένα ή εξειδικευμένα θέματα στην ευρύτερη συνομιλία βιωσιμότητας.
- **Αναδυόμενες σχέσεις:** Η εμφάνιση νέων συνδέσεων μεταξύ των hashtag με την πάροδο του χρόνου μπορεί να σηματοδοτήσει αλλαγές στο διάλογο, αντανakλώντας γεγονότα του πραγματικού κόσμου, αλλαγές πολιτικές ή αλλαγές στο δημόσιο συμφέρον.

Χρησιμοποίηση

Για τους ερευνητές και τους συμμετέχοντες στην εκστρατεία, το γράφημα δικτύου προσφέρει ένα στρατηγικό εργαλείο για να κατανοήσουν πώς συνυφαίνονται διαφορετικά θέματα βιωσιμότητας στον δημόσιο διάλογο. Βοηθά στον εντοπισμό των περιοχών που προσελκύουν την προσοχή και πώς συζητούνται διαφορετικά θέματα βιωσιμότητας μεταξύ τους.

7.7. Συχνότητα λέξεων έναντι συναισθήματος



Σχήμα 7-5: Συχνή χρήση λέξεων έναντι συναισθήματος

Περιγραφή

Το γράφημα Συχνότητα Λέξεων εναντίον Συναισθήματος είναι μια αναλυτική αναπαράσταση που αντιπροσωπεύει τη συχνότητα των λέξεων που εμφανίζονται στα tweets με τις σχετικές βαθμολογίες συναισθήματος. Αυτός ο τύπος οπτικοποίησης μπορεί να δομηθεί ως διάγραμμα διασποράς με συχνότητα λέξεων στον έναν άξονα και την βαθμολογία συναισθήματος στον άλλο, αποκαλύπτοντας μοτίβα σχετικά με το πώς ορισμένα σημεία συσχετίζονται με θετικά, ουδέτερα ή αρνητικά συναισθήματα στο πλαίσιο των συζητήσεων για τα απορρίμματα τροφίμων.

Λεπτομέρειες Οπτικοποίησης

Κάθε σημείο στο διάγραμμα διασποράς αντιπροσωπεύει μια μοναδική λέξη, με τη θέση της να καθορίζεται από τον αριθμό των φορών που εμφανίζεται στο σύνολο δεδομένων (συχνότητα) και τη μέση βαθμολογία συναισθήματος των tweets στα οποία εμφανίζεται. Λέξεις που εμφανίζονται συχνά σε tweets με υψηλές βαθμολογίες θετικού συναισθήματος μπορούν να απεικονίζονται στο πάνω δεξί τεταρτημόριο, υποδεικνύοντας μια ισχυρή συσχέτιση με θετικό λόγο, ενώ οι λέξεις που εμφανίζονται συχνά σε αρνητικά συμφραζόμενα μπορούν να βρεθούν στο κάτω δεξί τεταρτημόριο.

Ερμηνεία

- **Θετικές συσχετίσεις:** Λέξεις που αναφέρονται κυρίως σε θετικό πλαίσιο θα μπορούσαν να είναι ενδεικτικές συσκευές πρωτοβουλιών, εκστρατειών ευαισθητοποίησης ή λύσεων που σχετίζονται με τη σπατάλη τροφίμων.
- **Αρνητικές συσχετίσεις:** Αντίθετα, λέξεις που σχετίζονται με βαθμολογίες αρνητικών συναισθημάτων μπορούν να υπογραμμίσουν περιοχές ανησυχίας, προκλήσεις ή αμφιλεγόμενα θέματα που προκαλούν αρνητική αντίδραση του κοινού.
- **Ουδέτεροι όροι:** Λέξεις που συχνά αλλά έχουν βαθμολογία ουδέτερου συναισθήματος θα μπορούσαν να αντιπροσωπεύουν κοινούς όρους ή γενικές έννοιες στο θέμα της σπατάλης τροφίμων που δεν προκαλούν έντονα συναισθήματα.

Χρησιμοποίηση

Για τους φορείς επικοινωνίας και τους υπεύθυνους χάραξης πολιτικής, αυτό το γράφημα προσφέρει καθοδήγηση σχετικά με τη γλώσσα και την ορολογία που έχουν θετική ή αρνητική απάντηση στο κοινό. Μπορεί να ενημερώσει τον στρατηγικό σχεδιασμό των επικοινωνιακών εκστρατειών, επιτρέποντας τη δημιουργία μηνυμάτων που είναι πιο πιθανό να εμπλέξει το κοινό σε έναν εποικοδομητικό διάλογο σχετικά με τη σπατάλη τροφίμων.

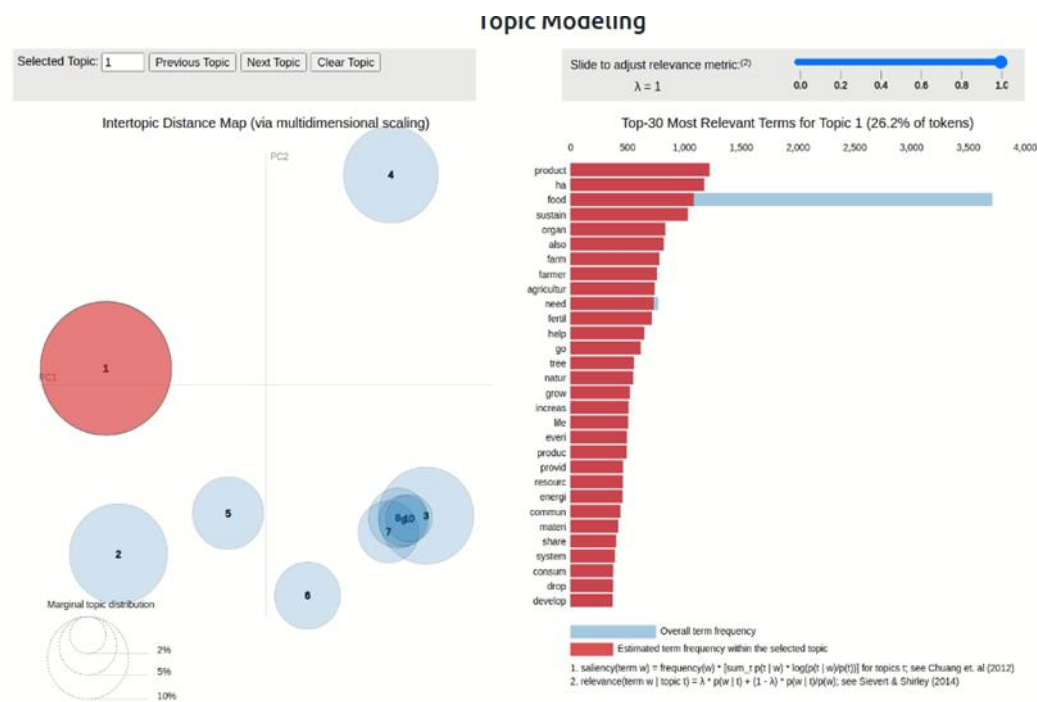
7.8. Μοντελοποίηση Θεμάτων

Περιγραφή

Τα κύρια στοιχεία των οπτικοποιήσεων ryLDAnis περιλαμβάνουν:

- **Διατοπικός χάρτης απόστασης (αριστερή πλευρά)**
 - Αυτό είναι ένα δισδιάστατο διάγραμμα διασποράς όπου κάθε συννεφάκι αντιπροσωπεύει ένα θέμα. Η απόσταση μεταξύ των φυσαλίδων υποδηλώνει την ομοιότητα μεταξύ των θεμάτων. Τα πιο κοντινά θέματα είναι πιο παρόμοια.
 - Το μέγεθος της φυσαλίδας αντιπροσωπεύει την επικράτηση του θέματος στο σώμα.
 - Αυτή η γραφική παράσταση βασίζεται συχνά στην Ανάλυση Κύριων Στοιχείων (PCA) ή στην πολυδιάστατη κλιμάκωση για τη μείωση των διαστάσεων των κατανομών όρων θέματος.
- **Σχετικότητα επίκαιρων όρων (Δεξιά πλευρά)**
 - Όταν επιλέγεται ένα θέμα κάνοντας κλικ σε ένα από τα συννεφάκια, αυτή η περιοχή εμφανίζει τους όρους που σχετίζονται περισσότερο με το επιλεγμένο θέμα.
 - Ένα ραβδόγραμμα δείχνει τους όρους που είναι πιο σχετικοί για το επιλεγμένο θέμα. Το μήκος της γραμμής αντιπροσωπεύει τον όρο συχνότητα εντός του θέματος.

Υπάρχει συνήθως ένα ρυθμιστικό (λ) που προσαρμόζει τη μέτρηση συνάφειας του όρου. Όταν το λ είναι κοντά στο 1, οι όροι που εμφανίζονται είναι οι πιο αποκλειστικοί για το θέμα. Όταν το λ είναι κοντά στο 0, οι όροι που εμφανίζονται είναι αυτοί που χρησιμοποιούνται περισσότερο στο θέμα, ανεξάρτητα από την αποκλειστικότητα.



Σχήμα 7-6: Μοντελοποίηση Θεμάτων

Η τοποθέτηση του ποντικιού πάνω από ένα συννεφάκι θα εμφανίσει πρόσθετες πληροφορίες σχετικά με το θέμα, όπως τον αριθμό του θέματος και τη συνολική συχνότητά του στο σώμα, ενώ η τοποθέτηση του δείκτη του ποντικιού πάνω από έναν όρο στο γράφημα ράβδων στα δεξιά θα επισημάνει τον τρόπο με τον οποίο αυτός ο όρος κατανέμεται σε διάφορα θέματα, όπως φαίνεται από τις μικρότερες γραμμές στην αριστερή γραφική παράσταση.

Συμπερασματικά, η ανάλυση του πίνακα ελέγχου αποκαλύπτει μια πολύπλευρη άποψη του δημόσιου λόγου σχετικά με τη σπατάλη τροφίμων και τη βιωσιμότητα. Η κατανομή του συναισθήματος δείχνει μια κυριαρχία θετικής δέσμευσης, υποδηλώνοντας μια αισιόδοξη προοπτική για πρωτοβουλίες και συζητήσεις σε αυτόν τον τομέα. Η ανάλυση των hashtag και η συνύπαρξή τους αποκάλυψε τα βασικά θέματα και τη διασύνδεση των θεμάτων στη συζήτηση για τη βιωσιμότητα.

Η τάση συναισθήματος χρονοσειράς δείχνει ένα ανταποκρινόμενο κοινό που αντιδρά σε γεγονότα και ειδήσεις, ενώ η εξέταση της συχνότητας των λέξεων έναντι του συναισθήματος υπογραμμίζει τη συγκεκριμένη γλώσσα που έχει απαίτηση στους χρήστες. Αυτή η λεπτή κατανόηση της δυναμικής της συνομιλίας προσφέρει σε οργανισμούς, ακτιβιστές και υπεύθυνους χάραξης πολιτικών πολύτιμες γνώσεις για το δημόσιο αίσθημα, δίνουν τη δυνατότητα να προσαρμόσουν αποτελεσματικά τις στρατηγικές και τις επικοινωνίες τους.

Το dashboard χρησιμεύει ως ένα ζωτικό εργαλείο για την παρακολούθηση και την κατανόηση της εξελισσόμενης αφήγησης γύρω από τη βιωσιμότητα στα μέσα κοινωνικής δικτύωσης. Αξιολογώντας τις γνώσεις, οι ενδιαφερόμενοι μπορούν να συμβάλουν σε έναν πιο

ενημερωμένο και αφοσιωμένο δημόσιο διάλογο, προωθώντας μια στροφή προς βιώσιμες πρακτικές και μειώνοντας τη σπατάλη τροφίμων σε παγκόσμια κλίμακα.

Παρόλες τις προκλήσεις, τα βασικά επιτεύγματα του προγράμματος αντικατοπτρίζουν σημαντική πρόοδο στον τομέα της τεχνολογίας και της ανάλυσης δεδομένων. Κατ' αρχάς, η αποτελεσματική συλλογή και επεξεργασία δεδομένων επιτεύχθηκε μέσω αυτοματοποιημένων σεναρίων και της χρήσης της MongoDB, η οποία επιτρέπει τον ισχυρό χειρισμό μεγάλων δεδομένων, ειδικά από τα μέσα κοινωνικής δικτύωσης. Επιπλέον, χρησιμοποιήθηκε προηγμένη ανάλυση δεδομένων, όπως η ανάλυση συναισθήματος, η ανάλυση συχνότητας και η μοντελοποίηση θεμάτων, για την εξαγωγή σημαντικών πληροφοριών από τα δεδομένα. Τέλος, ο διαδραστικός πίνακας εργαλείων που δημιουργήθηκε με τη χρήση του Dash Plotly επιτρέπει την διαισθητική και σε πραγματικό χρόνο απεικόνιση των δεδομένων, κάνοντας την ανάλυση πιο προσβάσιμη σε ένα ευρύ κοινό.

Ο αντίκτυπος του αναπτυχθέντος συστήματος φαίνεται να έχει πολλαπλές διαστάσεις. Πρώτον, αναμένεται να συμβάλει στην αυξημένη ευαισθητοποίηση κοινού και πολιτικών φορέων για τη σπατάλη τροφίμων, φέρνοντας στο προσκήνιο σημαντικά θέματα και συναισθήματα και βελτιώνοντας την κατανόηση των δημόσιων συζητήσεων γύρω από αυτό το ζήτημα. Επιπλέον, το σύστημα προσφέρει σημαντικές γνώσεις βασισμένες σε πληροφορίες και δεδομένα, κάτι που είναι πολύτιμο για τους ενδιαφερόμενους φορείς, όπως οι πολιτικοί, οι περιβαλλοντολόγοι και οι επιχειρηματίες, καθώς ενημερώνει τις στρατηγικές και πρωτοβουλίες τους για τη μείωση της σπατάλης τροφίμων. Τέλος, το αναπτυχθέν σύστημα λειτουργεί ως τεχνολογικό σημείο αναφοράς, καθώς δημιουργεί προηγούμενο για τη χρήση αναλύσεων από τα μέσα κοινωνικής δικτύωσης σε περιβαλλοντικά και κοινωνικά ζητήματα, επιδεικνύοντας τις δυνατότητες των μεγάλων δεδομένων στην προώθηση της αλλαγής.

Τέλος, σε ότι έχει να κάνει με μελλοντικές προοπτικές και επεκτάσεις του αναπτυχθέντος συστήματος, αναφέρονται οι κάτωθι:

- **Επεκτασιμότητα και Προσαρμογή:** Ο αρθρωτός σχεδιασμός του αναπτυχθέντος συστήματος επιτρέπει την εύκολη επεκτασιμότητα και την προσαρμογή σε άλλα κοινωνικά και περιβαλλοντικά ζητήματα.
- **Συνεχής Βελτίωση:** Υπάρχει δυνατότητα περαιτέρω βελτίωσης των τεχνικών ανάλυσης δεδομένων και ενσωμάτωσης πιο προηγμένων μεθόδων τεχνητής νοημοσύνης και μηχανικής εκμάθησης.
- **Ευρύτερη εφαρμογή:** Η μεθοδολογία και οι γνώσεις που αποκτήθηκαν, μπορούν να επεκταθούν σε άλλους τομείς, προσφέροντας ένα πρότυπο για την αξιοποίηση των δεδομένων κοινωνικών μέσων για κοινωνικό καλό.

Συμπερασματικά, το αναπτυχθέν σύστημα αποτελεί απόδειξη της δύναμης της τεχνολογίας στην κατανόηση και την αντιμετώπιση κρίσιμων κοινωνικών ζητημάτων. Γεφυρώνοντας το χάσμα μεταξύ του λόγου των μέσων κοινωνικής δικτύωσης και των πρακτικών γνώσεων,

συμβάλλει σημαντικά στην παγκόσμια προσπάθεια για τον μετριασμό της σπατάλης τροφίμων, οδηγώντας τελικά σε ένα πιο βιώσιμο και περιβαλλοντικά ενσυνείδητο μέλλον.

8. Συμπεράσματα

Κάθε πρόκληση που αντιμετωπίστηκε κατά τη διάρκεια των εργασιών ανάπτυξης του συστήματος παρείχε μια ευκαιρία για καινοτόμες λύσεις και μάθηση. Η επιτυχής αντιμετώπιση αυτών των προκλήσεων συνέβαλε στην ευρωστία και την αποτελεσματικότητα του συνολικού συστήματος, αποδεικνύοντας την προσαρμοστικότητα και την ανθεκτικότητα του συστήματος σε ένα δυναμικό περιβάλλον δεδομένων. Στον Πίνακα που ακολουθεί, συνοψίζονται οι εν λόγω προκλήσεις καθώς και οι λύσεις που δόθηκαν.

Πρόκληση	Περιγραφή	Λύση
Όγκος και πολυπλοκότητα δεδομένων	Ο χειρισμός μεγάλου όγκου δεδομένων από πλατφόρμες κοινωνικών μέσων όπως το Twitter και το Reddit παρουσίασε σημαντικές προκλήσεις όσον αφορά την αποθήκευση, την επεξεργασία και την ανάλυση.	Η εφαρμογή MongoDB για αποτελεσματική αποθήκευση και ανάκτηση δεδομένων και η χρήση των ισχυρών βιβλιοθηκών της Python για επεξεργασία και ανάλυση δεδομένων βοήθησε στη διαχείριση της πολυπλοκότητας και του όγκου των δεδομένων.
Επεξεργασία δεδομένων σε πραγματικό χρόνο	Η επεξεργασία και η ανάλυση δεδομένων σε πραγματικό χρόνο για την παροχή ενημερωμένων πληροφοριών ήταν μια πολύπλοκη εργασία.	Χρήση σεναρίων Python με προγραμματισμένες εργασίες (cron jobs) για τακτική συλλογή και επεξεργασία δεδομένων. Αυτό εξασφάλισε ότι ο πίνακας εργαλείων εμφανίζει τα πιο πρόσφατα δεδομένα χωρίς χειροκίνητη παρέμβαση
Ποιότητα δεδομένων και προεπεξεργασία	Η διασφάλιση της ποιότητας των δεδομένων που συλλέγονται από διάφορες πηγές και η τυποποίησή τους για ανάλυση ήταν πρόκληση.	Ανάπτυξη μιας ισχυρής μονάδας προεπεξεργασίας δεδομένων που περιλάμβανε κανονικοποίηση κειμένου, διακοπή αφαίρεσης λέξεων και άλλες τεχνικές NLP για τον καθαρισμό και την τυποποίηση των δεδομένων.

Ακρίβεια ανάλυσης συναισθήματος	Επίτευξη υψηλής ακρίβειας στην ανάλυση συναισθημάτων, ειδικά με τις αποχρώσεις και το πλαίσιο της γλώσσας των μέσων κοινωνικής δικτύωσης.	Η εφαρμογή του VADER, ενός εργαλείου ανάλυσης συναισθημάτων που είναι ειδικά προσαρμοσμένο στα συναισθήματα που εκφράζονται στα μέσα κοινωνικής δικτύωσης, βελτίωσε την ακρίβεια της ανάλυσης.
Αποτελεσματική Οπτικοποίηση Δεδομένων	Δημιουργία ενός διαισθητικού και διαδραστικού πίνακα εργαλείων που μπορεί να χειριστεί μεγάλα σύνολα δεδομένων χωρίς καθυστερήσεις απόδοσης.	Χρήση Dash Plotly για την κατασκευή του dashboard, που επέτρεψε την αποτελεσματική οπτικοποίηση δεδομένων. Η εφαρμογή στρατηγικών προσωρινής αποθήκευσης βελτίωσε περαιτέρω την απόδοση του πίνακα εργαλείων.
Πολυπλοκότητα Μοντελοποίησης Θεμάτων	Η μοντελοποίηση θεμάτων με μεγάλα σύνολα δεδομένων ήταν υπολογιστικά εντατική και πολύπλοκη.	Χρησιμοποιώντας το Gensim για τη μοντελοποίηση θεμάτων LDA και το PyLDAvis για διαδραστική απεικόνιση, που απλοποίησε την πολυπλοκότητα και βελτίωσε την ερμηνευτικότητα των αποτελεσμάτων.
Επεκτασιμότητα και Συντήρηση	Διασφάλιση της κλιμάκωσης του συστήματος με αυξανόμενα δεδομένα και ότι ήταν εύκολο να διατηρηθεί.	Ο σχεδιασμός μιας αρθρωτής αρχιτεκτονικής και η χρήση επεκτάσιμων τεχνολογιών όπως η MongoDB και η Python εξασφάλισε ότι το σύστημα θα μπορούσε να χειριστεί την ανάπτυξη και ότι ήταν διατηρητέο.

**Αντιγραφή
δεδομένων**

Η αποφυγή αντιγραφής δεδομένων από πολλαπλές πηγές ήταν ένα σημαντικό ζήτημα.

Εφαρμογή ελέγχων εντός της ενότητας συλλογής δεδομένων για τον εντοπισμό και τον αποκλεισμό διπλών εγγραφών που βασίζονται σε μοναδικά αναγνωριστικά.

9. Βιβλιογραφικές Αναφορές

1. Twitter API: Για συλλογή tweets που σχετίζονται με σπατάλη τροφίμων. [Τεκμηρίωση προγραμματιστή Twitter](<https://developer.twitter.com/en/docs>)
2. Reddit API: Χρησιμοποιείται για την απόξεση αναρτήσεων στο Reddit σχετικά με θέματα σχετικά με τα απόβλητα τροφίμων. [Τεκμηρίωση Reddit API](<https://www.reddit.com/dev/api>)
3. MongoDB: Βάση δεδομένων που χρησιμοποιείται για την αποθήκευση και τη διαχείριση συλλεγόμενων δεδομένων. [Επίσημος ιστότοπος MongoDB](<https://www.mongodb.com/>)
4. Dash Plotly : Για δημιουργία διαδραστικών πινάκων εργαλείων για οπτικοποίηση δεδομένων. [Dash Plotly Τεκμηρίωση](<https://dash.plotly.com/>)
5. Γλώσσα Προγραμματισμού Python : Η κύρια γλώσσα που χρησιμοποιείται για τη δημιουργία σεναρίων και την επεξεργασία δεδομένων. [Επίσημος ιστότοπος Python](<https://www.python.org/>)
6. NLTK (Natural Language Toolkit): Μια βιβλιοθήκη Python που χρησιμοποιείται για εργασίες επεξεργασίας φυσικής γλώσσας. [Τεκμηρίωση NLTK](<https://www.nltk.org/>)
7. Gensim : Βιβλιοθήκη Python που χρησιμοποιείται για μοντελοποίηση θεμάτων και άλλες εργασίες NLP. [Gensim Τεκμηρίωση](<https://radimrehurek.com/gensim/>)
8. PyLDAvis : Μια βιβλιοθήκη Python για διαδραστική οπτικοποίηση μοντέλων θεμάτων. [PyLDAvis GitHub Repository](<https://github.com/bmabey/pyLDAvis>)
9. VADER Sentiment Analysis: Ένα εργαλείο ανάλυσης συναισθήματος βασισμένο σε λεξικά και κανόνες, ειδικά προσαρμοσμένο στα συναισθήματα που εκφράζονται στα μέσα κοινωνικής δικτύωσης. [VADER GitHub Repository](<https://github.com/cjhutto/vaderSentiment>)
10. NetworkX : Πακέτο Python για τη δημιουργία, το χειρισμό και τη μελέτη της δομής, της δυναμικής και των λειτουργιών πολύπλοκων δικτύων. [ΔίκτυοX Τεκμηρίωση](<https://networkx.org/>)
11. Plotly : Μια βιβλιοθήκη γραφημάτων για τη δημιουργία διαδραστικών γραφημάτων ποιότητας δημοσίευσης στο διαδίκτυο. [Plotly Python Graphing Library](<https://plotly.com/python/>)
12. Pandas: Ένα γρήγορο, ισχυρό, ευέλικτο και εύχρηστο εργαλείο ανάλυσης και χειρισμού δεδομένων ανοιχτού κώδικα. [Τεκμηρίωση Pandas](<https://pandas.pydata.org/>)
13. Cron: Ένας προγραμματιστής εργασιών με βάση το χρόνο σε λειτουργικά συστήματα υπολογιστών παρόμοια με το Unix. [Σελίδα Cron Wikipedia](<https://en.wikipedia.org/wiki/Cron>)
14. M. Steyvers and T. Griffiths, "Probabilistic topic models," in Handbook of latent semantic analysis, 2007, pp. 424-440.
15. D. M. Blei, Y. A. Ng and M. I. Jordan, "Latent Dirichlet allocation," Journal of Machine Learning Research 3, pp. 993-1022, 2003.

16. D. M. Blei, "Probabilistic topic models", *Communications of the ACM*, vol. 55, no. 4, pp. 77-84, 2012.
17. Xingyu Wang, Lida Zhang, Diego Klabjan, "Keyword-based Topic Modeling and Keyword Selection", 2021 IEEE International Conference on Big Data (Big Data), pp 1148-1154.
18. Amna Meddeb, Lotfi Ben Romdhane, "Using Topic Modeling and Word Embedding for Topic Extraction in Twitter", 26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2022)