

Title: Hot Topic detection and tracking on AFCON 2021 in Cameroon

(The french version is after the english one)



THE BACKGROUND:

Nowadays, many people turn to express their feelings regarding events through social media, thus generating a vast corpus of conversations on a vast range of topics. Hot topic detection being an important knowledge discovery task on social data streams turns out to be very useful in such a context. They help to determine hot topics that are being discussed the most on a social platform. One of the main challenges of this task is the processing of a very large social dataset in a sequence order efficiently and effectively.

The Total Energies 2021 Africa Cup of Nations (CAN Total Energies 2021) took the continent by storm for a good 29 days, with the recently concluded 33rd edition acclaimed as the most widely followed in the history of the competition. #AFCON2021 engendered a surplus of feisty content, laden with agenda-setting, crescendoing the continent into a fever pitch. With mainstream and social media across and beyond the continent awash with technical analyses of the sport. This exciting event was about 24 countries, sparring through 52 matches and ending up with a Senegal-Egypt finale with the Teranga Lions emerging as first-time champions! With Cameroon gleaning 3rd place after a spectacular Lion-montada against Burkina Faso. Senegal's Sadio Mané and Edouard Mendy winning best player and

goalkeeper, respectively, Cameroon's Vincent Aboubakar finishing as top goalscorer with 8 goals, with Karl Toko Ekambi (5 goals) in tow, Cameroon's "CAN sucrée" thrilled spectators to the very end. However, the murmuring ranged from debatable concerns about the refereeing and CAF COVID-19 regulations to improbable claims about accommodation. But it wasn't long before it morphed into a berserk attack on the host country, more could be read here.

THE PROBLEM:

The continuously increasing artifacts about the AFCON2021 on social media make it difficult to quickly gain insights into certain themes that best describe the noise we had. To facilitate such comprehension, and to provide a better understanding of the noise around the AFCON2021 this 4 week's challenge will propose a model to discover the hot topics and patterns of topics evolutions in the AFCON2021. The determination of these hot topics could help give a better picture of the main concerns of fans as well as what they did like the most, thus helping in improving next organizations and making an even better AFCON for the coming years.

THE PROJECT GOALS:

1. Collect and pre-process the historical interactions data of (Facebook, Twitter, ...) users for the months July 2021 to February 2022;
2. Collect, pre-process and analysis of the frequency of hot terms (events, words, publications, ...) related to AFCON 2021: extract hot terms and assignation of weight. Hot terms are extracted based on three features: the frequency of the terms used in the document collection; the location of the terms within a document; the breadth of terms distribution in the document collection. ref(4);
3. Provide a Hot Term Recognition model with Timeline Analysis
4. Multidimensional Sentence Modeling
5. Evaluation and validation of model
6. Provide an interactive dashboard about results

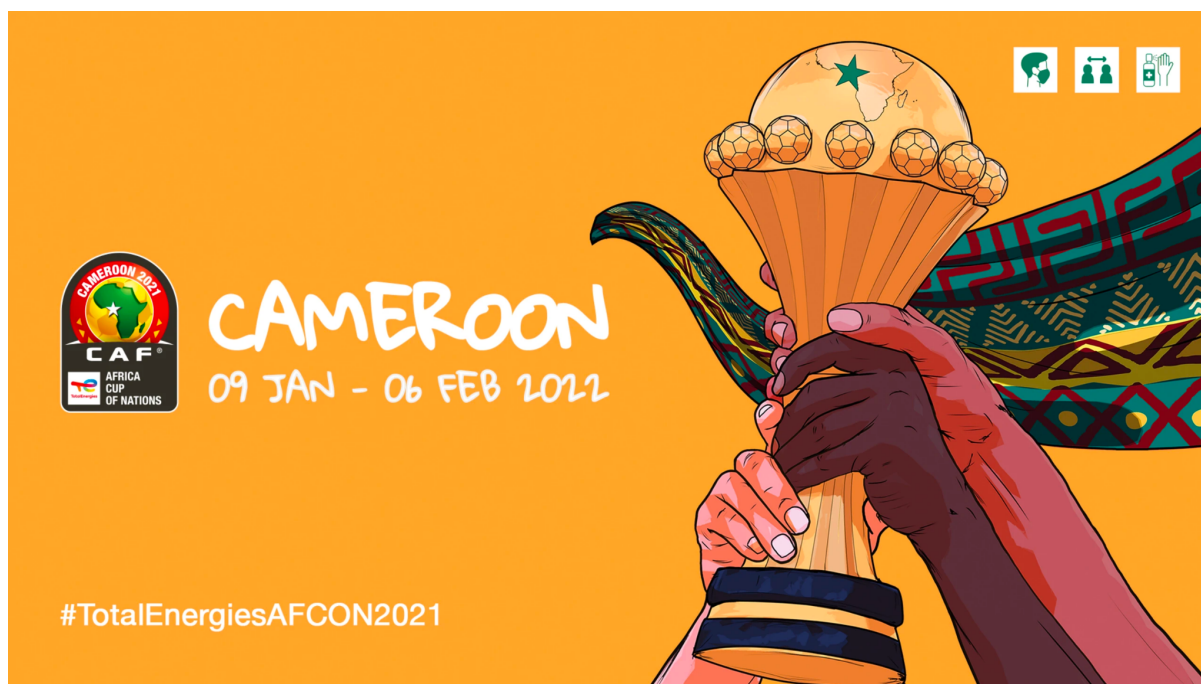
THE LEARNING OUTCOMES:

1. Social Media Data Extraction
2. Data pre-processing
3. Exploratory Data Analysis
4. Transformers and Graph-based models
5. Interactive Dashboard

THE TASKS & TIMELINE:

Week 1	Week 2	Week 3	Week 4
Tools preparation for data extraction + Data pre-processing	Analysis + Model construction and Training	Model construction and training +Dashboard	Preparation of the report and dashboard
Output: Data	Output: Data visualization, Hot topics	Output: Hot topics, Dashboard	Output: Presentation of the report and dashboard

Titre : Détection et suivi des sujets chauds sur la CAN2021 au Cameroun



CONTEXTE :

De nos jours, de nombreuses personnes expriment leurs sentiments sur des événements par le biais des médias sociaux, ce qui génère un vaste corpus de conversations sur un large éventail de sujets. La détection des sujets chauds, qui est une tâche importante de découverte de connaissances sur les flux de données sociales, s'avère très utile dans un tel contexte. Elle permet de déterminer les sujets chauds qui sont le plus discutés sur une plateforme sociale. L'un des principaux défis de cette tâche est le traitement d'un très grand ensemble de données sociales dans un ordre séquentiel de manière efficace et effective.

La Coupe d'Afrique des Nations Total Energies 2021 (CAN Total Energies 2021) a pris le continent d'assaut pendant 29 jours, la 33e édition qui vient de s'achever étant acclamée comme la plus suivie de l'histoire de la compétition. La #CAN2021 a engendré un surplus de contenu fougueux, chargé d'agenda, qui a fait monter la fièvre sur le continent. Les médias grand public et sociaux, à travers et au-delà du continent, ont été inondés d'analyses techniques du sport. Cet événement passionnant a rassemblé 24 pays, qui se sont affrontés au cours de 52 matchs, pour se terminer par une finale Sénégal-Égypte où les Lions de la

Téranga ont remporté leur premier titre ! Le Cameroun, quant à lui, a décroché la troisième place après une spectaculaire Lion-montada contre le Burkina Faso. Les Sénégalais Sadio Mané et Edouard Mendy, respectivement meilleur joueur et meilleur gardien de but, le Camerounais Vincent Aboubakar, meilleur buteur avec 8 buts, et Karl Toko Ekambi (5 buts), la "CAN sucrée" du Cameroun a enthousiasmé les spectateurs jusqu'au bout. Pourtant, les murmures allaient de préoccupations discutables sur l'arbitrage et le règlement CAF COVID-19 à des revendications improbables sur l'hébergement. Mais il n'a pas fallu longtemps pour qu'elle se transforme en une attaque en règle contre le pays hôte, en savoir plus ici.

LE PROBLÈME :

L'augmentation continue des artefacts concernant la CAN2021 sur les médias sociaux rend difficile la compréhension rapide de certains thèmes qui décrivent le mieux le bruit que nous avons eu. Pour faciliter une telle compréhension, et pour fournir une meilleure compréhension du bruit autour de la CAN2021, ce défi de 4 semaines proposera un modèle pour découvrir les sujets chauds et les modèles d'évolution des sujets de la CAN2021. La détermination de ces sujets chauds pourrait aider à donner une meilleure image des principales préoccupations des fans ainsi que de ce qu'ils ont le plus apprécié, aidant ainsi à améliorer les prochaines organisations et à faire une CAN encore meilleure pour les années à venir.

LES OBJECTIFS DU PROJET :

1. Collecter et prétraiter les données historiques d'interactions des utilisateurs (Facebook, Twitter, ...) pour les mois de juillet 2021 à février 2022 ;
2. Collecte, pré-traitement et analyse de la fréquence des termes chauds (événements, mots, publications, ...) relatifs à la CAN 2021 : extraction des termes chauds et attribution d'un poids. Les termes chauds sont extraits sur la base de trois caractéristiques : la fréquence des termes utilisés dans la collection de documents ; l'emplacement des termes dans un document ; l'ampleur de la distribution des termes dans la collection de documents. ref(4) ;
3. Fournir un modèle de reconnaissance des termes chauds avec une analyse chronologique

4. Modélisation multidimensionnelle des phrases
5. Évaluation et validation du modèle
6. Fournir un tableau de bord interactif sur les résultats

LES RÉSULTATS D'APPRENTISSAGE :

1. Extraction de données sur les médias sociaux
2. Prétraitement des données
3. Analyse exploratoire des données
4. Transformateurs et modèles basés sur des graphiques
5. Tableau de bord interactif

LES TÂCHES ET LE CALENDRIER :

Semaine 1	Semaine 2	Semaine 3	Semaine 4
Préparation des outils pour l'extraction des données + Prétraitement des données	Analyse + Construction de modèles et Training	Construction de modèles et Training +Tableau de bord	Préparation du rapport et du tableau de bord
Sortie : Données	Sortie : Visualisation des données, sujets chauds	Sortie : Sujets chauds, tableau de bord	Sortie : Présentation du rapport et du tableau de bord

Refs:

1. <https://cs.emis.de/LNI/Proceedings/Proceedings200/165.pdf>
2. <https://people.cs.pitt.edu/~chang/265/proj10/sisref/8.pdf>
3. <https://ieeexplore.ieee.org/document/8713730>
4. <https://slejournal.springeropen.com/track/pdf/10.1186/s40561-015-0019-6.pdf>
5. <https://hrcak.srce.hr/file/325826>
6. <https://www.afm-marketing.org/sites/default/files/atoms/files/Hot-topics-marketing-afm2018.pdf>