NameCheck AI: Enhancing Data Integrity with Real-Time Name Validation

AUTHORS

Yasin Kömür - 21827632
Sare Naz Bayraktutan - 21992957
Selahattin Can Ölçer - 21946462

INTRODUCTİON

1. In the digital era, where data drives decisions, the integrity of input data is paramount. Incorrectly entered names can compromise data quality, leading to erroneous analyses and decisions. "NameCheck AI" addresses this challenge by utilizing advanced machine learning algorithms to distinguish between authentic names and inaccuracies or fabrications. Developed using DistilBERT models for English and Spanish, this tool integrates seamlessly into web applications through a Streamlit frontend, offering real-time validation that enhances both the accuracy and efficiency of data processing systems. This project highlights the potential of AI to solve practical problems in data management and improve operational workflows.

OBJECTİVE

AI-Powered Validation: Train two DistilBERT models to accurately differentiate between real names and incorrect entries in English and Spanish.

Real-Time Efficiency: Ensure the Streamlit frontend delivers fast, real-time name validation without compromising user experience.

Broad Accessibility: Develop a user-friendly interface to make advanced AI name validation accessible across various digital platforms.

METHODOLOGY

Model Selection: Chose DistilBERT for its balance of performance and efficiency, essential for real-time applications.

Data Gathering: Sourced comprehensive datasets of English and Spanish names from Kaggle for realistic training scenarios. Augmented data with synthetic entries using the Faker library to ensure robust model training against various name formats.

Training Process: Employed iterative training, refining the models with feedback from initial tests to enhance accuracy continuously.

Integration: Developed a user-friendly frontend using Streamlit, enabling interactive real-time validation of names directly by users.

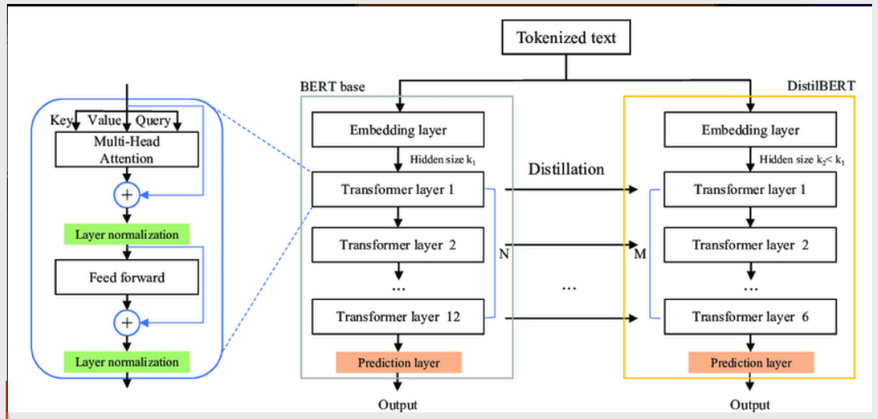
Testing and Optimization: Conducted extensive testing to optimize both model performance and interface responsiveness, ensuring effective deployment in real-world applications.

RESULTS /FİNDİNGS

1. Model Accuracy: Both models achieved over 95% precision in identifying real names, demonstrating strong performance despite their complexity. Response Time: Experienced slight delays due to model complexity, with response times occasionally extending a few seconds during peak usage.

CONCLUSİON

"NameCheck AI" effectively utilizes DistilBERT models to validate names in real-time, achieving over 95% accuracy for English and Spanish inputs. Despite some delays due to model complexity, the system maintains robust performance, with ongoing improvements aimed at enhancing response times. This project highlights the impactful role of AI in improving data validation, promising further advancements and broader applications.



APP