REAL-TIME STRESS DETECTION SYSTEM USING PHYSIOLOGICAL SIGNALS

AN AI-BASED FRAMEWORK HARNESSING THE APACHE SPARK ENVIRONMENT

LUDOVICO GANDOLFI

WHY STRESS MONITORING?

In today's fast-paced and demanding world, stress
has become an ever-present reality for many individuals.

Stress can manifest in various forms and affect
our mental and physical well-being.

Stress monitoring enables individuals to gain as a preventive measure
against the negative consequences of chronic stress and potentially provides support for health professionals overseeing sensible patients.

The aim of this project is to develop a **real-time stress detection system** using physiological signals, specifically Heart Rate Variability (HRV) and Electrodermal Activity (EDA).
The system is designed to continuously monitor these signals and detect instances of stress in real-time. This serves a wide range of applications, including stress management, mental health monitoring, and intervention systems.

TABLE OF CONTENTS

01. DATA COLLECTION

This project makes use of the research materials (datasets) developed for a scientific study at the Aoyama Gakuin University of Tokyo, Japan. The materials have been collected in the process of defining the academic paper **Nkurikiyeyezu, Yokoubo and Lopez (2020)** "*The Effect of Person-Specific Biometrics in Improving Generic Stress Predictive Models*", Journal of Sensors and Materials.

As the project is largely based on the findings of this publication, the raw data requested some minor modifications, namely:

- Raw data has been preprocessed to **remove noise, artifacts, and irrelevant information**
- Feature extraction techniques have been applied to **extract relevant features** from the signals

After performing some exploratory data analysis, we identified our target variable "condition label", taking value 0, 1 or 2 for "Baseline", "Amusement" and "Stress" conditions respectively.
"Baseline" and "Amusement" conditions are then joined, leaving a binary response variable, taking value 0 and 1 for "NoStress" and "Stress" conditions respectively.

02. PREDICTIVE MODEL - OUR TOOLS

Spark DataFrames, **Spark Structured Streaming**, and **Apache Kafka** forms the backbone of our real-time stress detection pipeline.

Spark DataFrames enable efficient processing of physiological data, while Spark Structured Streaming ensures **fault-tolerant and scalable stream processing**. Apache Kafka provides scalable ingestion and messaging capabilities, ensuring reliable delivery of streaming data to the processing pipeline.

By leveraging this robust pipeline architecture, the stress detection system can analyze streaming physiological data in real-time, identify stress patterns, and trigger timely interventions or alerts to users based on their stress levels.
Using these tools ensures the pipeline architecture is scalable and adaptable, capable of **handling increasing volumes of streaming data** as the user base grows. It can also accommodate future enhancements and integrations with additional data sources or processing logic to enhance stress detection capabilities.

02. PREDICTIVE MODEL - MODEL SELECTION

Given the different nature of the two datasets, two distinct predictive models have been trained and validated. For each of them, a randomised **Train/Test split** has been performed and different classification models have been fitted using **Cross-Validation** to optimize the combination of hyperparameters yielding the best performing model.

Fitted models include Logistic Regression, SVM, Decision Trees, Random Forest and Gradient Boosted Trees.

Given the nature of the issue, the models have been **optimized for Recall**. The choice has been taken to minimise the occurrence of False Negatives (missed detection of a stress condition).

The following are the best performing models for each dataset:

Heart Rate Variability (HRV)

Selected model: **DecisionTreeClassifier**
Recall on Test set: **0.982**
F-1 Score on Test set: 0.985
Hyperparameters:

- maxDepth: 15
- maxBins: 20

Electrodermal Activity (EDA)

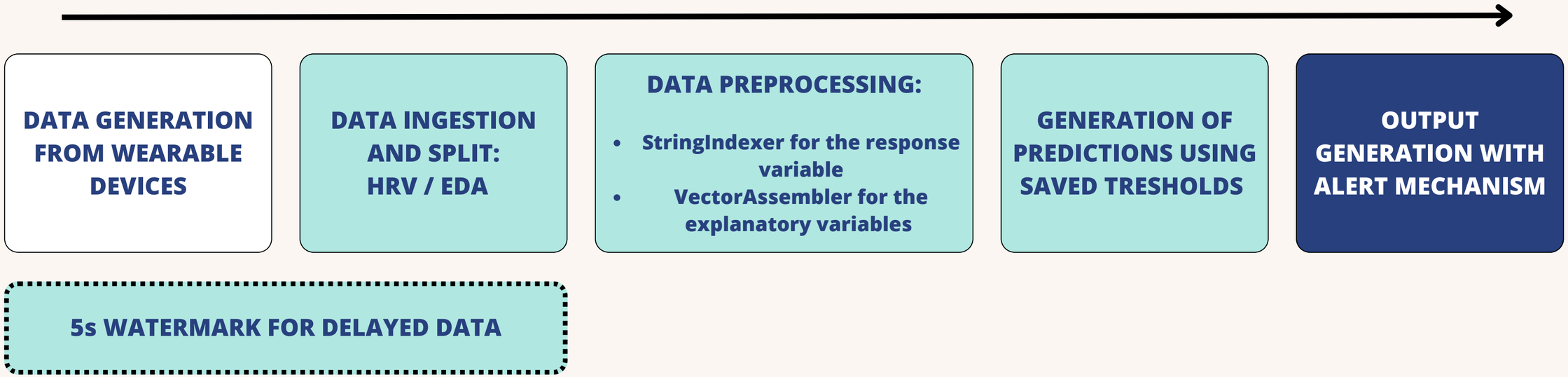
Selected model: **RandomForestClassifier**
Recall on Test set: **0.975**
F-1 Score on Test set: 0.976
Hyperparameters:

- numTrees: 20
- maxDepth: 5

03. STREAMING PIPELINE

Wearable devices send streaming physiological data through **Apache Kafka**, that provides a distributed publish-subscribe messaging system. Kafka acts here as a buffer, decoupling data ingestion from data processing, and ensuring data durability and reliability.

By integrating Apache Kafka with **Spark Structured Streaming**, the pipeline can ingest streaming physiological data from wearables in a **fault-tolerant** (a 5 seconds watermark is set) and scalable manner, enabling real-time stress detection with **low latency** and **high throughput**.



04. USE CASES - PERSONAL HEALTH & WELLBEING

A first use case pertains the application of the system in a private scenario, providing support for **personal stress management and lifestyle optimisation**:

- Individuals can use wearable devices equipped with stress detection technology to monitor their stress levels throughout the day. Real-time notifications could prompt users to engage in **stress-relief activities.** By fostering self-awareness and providing actionable insights, these devices empower individuals to **manage stress effectively** and maintain better overall well-being.

- Moreover, this technology could easily integrate with other health metrics such as physical activity, eating habits and sleep patterns to provide holistic insights into **users' lifestyle habits**. By correlating stress levels with factors such as exercise, nutrition, and social interactions, wearable devices can help users identify lifestyle changes that promote **stress reduction and resilience**.

04. USE CASES - MENTAL HEALTH MONITORING

This second use case pertains the application of the system in a clinical setting, providing support for **early intervention, treatment adherence and progress tracking of psychiatric patients**:

- Traditional methods of assessing stress in psychiatric wards rely heavily on subjective observations or self-reporting, which may not always be accurate. A real-time stress detection system offers an **objective and quantifiable measure of stress levels**, enhancing the accuracy of assessment and enabling more informed decision-making by healthcare providers. Early detection would allow healthcare providers to intervene promptly, **providing support and interventions before the stress escalates into a crisis situation**.

- For individuals receiving treatment for mental health conditions, wearable devices can assist in **monitoring treatment adherence and tracking symptom progression** over time. Moreover, timely adjustments to treatment plans can be made based on real-time data, leading to more **personalized** and effective **mental health care**.

05. DATA GOVERNANCE

The deployment of this framework carries several **data governance concerns** that should be addressed to ensure ethical, legal, and responsible use of the data. Here are some main issues to consider:

- **Privacy and Consent**
  Issue: Ensuring that users provide informed consent for the collection, storage, and processing of their physiological data is crucial. This includes transparently communicating the purpose of data collection, how the data will be used, and any potential risks or implications.
  Mitigation: Implement robust consent mechanisms, such as explicit opt-in/opt-out processes, privacy policies, and user agreements.

- **Data Security and Confidentiality**
  Issue: Physiological data is highly sensitive and must be protected from unauthorized access, misuse, and breaches.
  Mitigation: Implement strong data security measures, including encryption, access controls, dual-factor authentication mechanisms, and regular security audits.

Compliance with industry standards and regulations such as **GDPR**, **HIPAA**, or **CCPA** should be ensured to guarantee confidentiality and integrity of personal health data.

In summary, a real-time stress detection framework could hold tremendous potential in empowering individuals to manage stress effectively and supporting mental health monitoring and intervention. By leveraging wearable devices and advanced analytics, we can ultimately contribute to a healthier, happier, and more resilient society.

Thank you for reading.

Ludovico Gandolfi