

Università degli Studi dell'Insubria  
Dipartimento di Scienze Teoriche e Applicate (DiSTA)  
Corso di Laurea Triennale in Informatica



# “Applicazione di Tecniche di Machine Learning per la Determinazione dello Stato di Stress nei Conducenti”

Relatore: Prof. Silvia Corchs

Tesi di Laurea di  
Cristian Citterio  
Matricola 745297

Anno Accademico 2023-2024

## Contesto e Scopo della Ricerca

---

La ricerca si inserisce nel contesto della sicurezza stradale, concentrandosi sull'importanza della prevenzione degli incidenti causati dallo stress. L'obiettivo principale di questo progetto è sviluppare un modello predittivo in grado di identificare lo stress del conducente utilizzando una combinazione di segnali fisiologici e telemetrici.

## Metodologia e dataset di riferimento

---

La metodologia utilizzata si basa sull'utilizzo di un dataset multimodale che fornisce informazioni dettagliate sulle varie forme di distrazione alla guida. Il dataset utilizzato come riferimento principale è "A multimodal dataset for various forms of distracted driving", pubblicato su Scientific Data nel 2017. Attraverso l'uso di un simulatore di guida realistico e sensori non invasivi, sono stati raccolti dati sulle reazioni fisiologiche e comportamentali dei conducenti, nonché sulle loro prestazioni di guida.

Dopo una fase iniziale di pre-elaborazione e filtraggio dei dati, sono stati valutati gli stati di stress dei soggetti che hanno partecipato al simulatore di guida. Questa valutazione è stata effettuata utilizzando vari classificatori e tecniche di Machine Learning, che hanno permesso di analizzare e comprendere meglio l'impatto delle distrazioni sulla guida. Sono stati addestrati modelli predittivi utilizzando classificatori come k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Naive Bayes, Random Forest, Decision Tree e Logistic Regression. Inoltre, sono state esplorate tecniche di validazione incrociata come la *StratifiedShuffleSplit*, la *Five Fold Cross Validation* e la *Leave One Subject Out* per valutare le prestazioni dei modelli.

## Pre-elaborazione dei Dati

---

La pre-elaborazione dei dati è stata un'importante fase preliminare del processo di analisi. Sono state applicate tecniche di filtraggio per rimuovere il rumore dai segnali e ridurre l'effetto di interferenze esterne. Inoltre, sono state eseguite operazioni di normalizzazione per standardizzare i dati e ridurre le discrepanze tra le diverse scale di misura.

## Estrazione delle Features

---

Sono state estratte caratteristiche sia dalle serie temporali dei segnali fisiologici che dalle loro rappresentazioni spettrali. Inoltre, sono state calcolate statistiche descrittive come media, mediana, deviazione standard, ecc., per catturare ulteriori informazioni sulle distribuzioni dei segnali.

## Metodi di Validazione Utilizzati

---

Nelle fasi successive sono stati utilizzati diversi metodi di validazione incrociata, tra cui la 'Five Fold Cross Validation', la 'Leave One Out' e 'StratifiedShuffleSplit', per valutare le prestazioni del modello su nuovi dati.

## Risultati Principali Ottenuti: Classificazione Multiclasse e Binaria

---

In questo progetto, sono stati impiegati due approcci di classificazione distinti per determinare lo stato di stress dei conducenti durante la guida: la classificazione multiclasse e la classificazione binaria.

La classificazione multiclasse è stata utilizzata per distinguere quattro tipi di stress: non stress, stress cognitivo, emotivo e sensorimotorio.

D'altra parte, la classificazione a due etichette è stata impiegata per discriminare tra due stati distinti, *stress* versus *non stress*, semplificando la complessità della valutazione ad una decisione binaria.

Nella classificazione binaria, i risultati hanno dimostrato un'accuratezza generale superiore. Durante l'analisi dei dati, è emerso che per la tecnica di validazione Stratified-ShuffleSplit, i classificatori migliori sono stati l'SVM (ottimizzato con il *Gridsearch*) per la classificazione binaria ed il KNN (ottimizzato con il *Gridsearch*) per la classificazione multiclasse, che hanno ottenuto rispettivamente un'accuratezza del 94% e del 80% nel rilevare lo stato di stress. Per quanto riguarda l'utilizzo della 'Leave One Out' e della 'Five Fold Cross Validation', i risultati mostrano che l'ANN è stato il classificatore più performante, con un'accuratezza media del 80% per la classificazione a due etichette e del 55% per la classificazione a quattro etichette. Dalle valutazioni ottenute, la precisione totale nella classificazione multiclasse dello stress rappresenta un ambito che necessita di approfondimenti, indicando così la possibilità di ulteriori progressi.

## Utilizzo della PCA e del ROC

---

Nella presente ricerca, sono state utilizzate la Principal Component Analysis (PCA) e la Receiver Operating Characteristic (ROC) per valutare e migliorare le prestazioni dei modelli predittivi.

## Considerazioni sull'Applicabilità Pratica e Possibili Sviluppi Futuri

---

La ricerca è stata condotta utilizzando strumenti e risorse come Python e Scikit-learn, e potrebbe essere ottimizzata esplorando altri approcci e tecniche di machine learning, come le reti neurali profonde. È importante valutare l'efficacia del modello anche in situazioni reali di guida per confermarne l'utilità pratica. L'integrazione con sistemi avanzati di assistenza alla guida potrebbe migliorare ulteriormente la sicurezza stradale.

## Conclusioni e Implicazioni della Ricerca

---

Gli esperimenti indicano che l'uso di classificatori avanzati e segnali fisiologici può essere promettente per affrontare il problema dello stress del conducente, ma sono necessarie ulteriori ricerche per ottimizzare i modelli. Questo potrebbe influenzare positivamente la progettazione di sistemi di sicurezza nei veicoli, ma è importante considerare anche le implicazioni sociali ed etiche, come la privacy e la sicurezza dei dati personali, garantendo politiche e regolamenti appropriati.