**Proposte per project works MPA 2023-2024**

**Test n.01**

Si vuole costruire uno stimatore pe la torbidità osservata ina acqua a patie da dati satellitari. Viene fornito un file csv con coppie di torbidità – risposte spettrali rilevati da satellite. I dati hanno anche una colonna di validità sulla base della presenza di copertura nuvolosa. Lo stimatore deve essere valutato in temini di errore atteso e si desidera anche capire quali bande spettrali sono importanti per la previsione del dato di torbidità. Come confronto si possono usare formule semplici che includono singole bande spettrali.

**Test n.02**

Sui vuole studiare la dipendenza del consumo di una nave dalla velocità e da altri parametri misurati a bodo, al fine di valutare le prestazioni e l’impatto dei vai parametri. I dati forniti (in formato xlsx) sono però molto sbilanciati su alcune condizioni tipiche pe cui va posta attenzione alla loro rappresentatività. Anche la qualità dei dati non è perfetta e va valutata prima di costruire qualunque modello. Le colonne sono sicuramente ridondanti e vanno individuate quelle più utili per lo studio.

**Test n.03**

Un laboratorio di controllo qualità deve classificare campioni veto in diverse classi, utilizzando misure di spettrometria di massa. L’obiettivo, oltre che ottenere la risposta mediante un classificatore black-box, è anche capire se le classi differiscono per qualche caratteristica chimica o se possono esserci regole di tipo chimico pe inferire la classe in modo diretto.

**Test n.04**

Dato un dataset di immagini a livelli di grigio, si vuole costruire un classificatore neurale in grado di distinguere le immagini che contengono un veicolo al centro dell’inquadratura. Si suggerisce di partire da una rete già addestrata come estrattore di features su cui realizzare sono l’ultimo stadio (che potrebbe anche non essere neurale).

**Test n.05**

Si vuole costruire un detector di segnali stradali, utilizzando come base il segnale di stop. Vengono fornite immagini stradali con e senza il cartello in questione, ed un file di groundtruth con un rettangolo per ogni immagine di training. Si richiede di fare data augmentation ed eventualmente arricchire il dataset con altro materiale di qualunque provenienza. Si possono usare detector semplici addestrati da zero oppue detector già addestrati su cui fare solo fine tuning.

**Test n.06**

A partire da un campione di hard disks dei quali sono disponibili i dati SMART fino alla rottura. Si vuole capire quali dati sono utili per anticiparla e se è possibile realizzare un sistema di alert che segnali quali dischi si romperanno entro un tempo fissato. Il problema può essere impostato sia come classificazione binaria, sia come one-class classification (o anomaly detection).

**Test n.07**

Una RSA sta facendo uno studio sulla previsione di quali pazienti sono soggetti a lesioni da decubito sulla base di una ricca selezione di variabili rilevate all’atto del ricovero. La valutazione può essere binaria (con due livelli di soglia, LDP e LDPS) o multiclasse (NPUAP). Molte variabili sono probabilmente inutili, e diverse presentano mancanze che vanno gestite. Si vuole, se possibile, una formula semplice che gli operatori sanitari possano applicae al ricoveo per decidere il tipo di dotazioni da utilizzare.

**Test n.08**

Si vuole studiare la domanda di energia elettrica di una piccola città (prossima all’equatore, quindi in assenza di un chiaro ciclo stagionale), in funzione di parametri meteorologici, al fine di capire quale è il pattern nominale giornaliero di domanda e come le variabili meteo lo influenzano. Si può approcciare il problema sia come analisi di serie temporali sia come regressione statistica utilizzando il tempo come variabile indipendente.

**Test n.09**

Un laboratorio universitario sta facendo uno studio sulla depressione negli anziani (CDR) sulla base di un vasto insieme di variabili dedotte da interviste ed osservazioni. Lo scopo è prevedere quali pazienti sono a rischio, valutato in 6 classi. Visto lo sbilanciamento nel dataset, si possono ridurre le classi in modo da ottenere almeno una discriminazione su almeno 2-3 livelli.

**Test n.10**

Un ente locale monitora da anni la concentrazione in mare di una specifica alga tossica. Viene fornito in parallelo un vasto campione di dati meteo per la stessa zona. Si chiede di realizzare uno strumento previsionale che sulla base del meteo di un certo giorno (e dei giorni precedenti) stima la presenza di alga e soprattutto se il valore supera una soglia di attenzione per il potenziale pericolo (100.000). I biologi sanno che le condizioni meteo favorevoli pe la proliferazione (bloom) algale si osservano fino a 10 giorni pima dell’effettivo bloom.

**Test n.11**

Un ente locale monitora da anni la concentrazione in mare di una specifica alga tossica. Viene fornito un dataset di osservazioni spaziali distinte in N stazioni di misura. Si chiede di realizzare uno strumento previsionale che sulla base del meteo di un certo giorno in una certa stazione, stima la presenza di alga e soprattutto se il valore supera una soglia di attenzione per il potenziale pericolo (10.000). L’indicativo della stazione va inserito come variabile categorica. Interessa anche capire sei modello unificato performa più o meno bene a seconda (o addirittura senza) l’informazione di stazione.

**Test n.12**

Un anemometro regista dati di intensità e direzione media del vento a cadenza fissa di 1 ora. Si vuole valutare se esiste la possibilità di prevedere le direzioni di arrivo di vento forte in modo da orientare un generatore eolico nella direzione ottimale. La soluzione va comparata con quella a direzione fissa (da scegliere), supponendo che l’errore di angolo produca una perdita di efficienza proporzionale al seno dell’angolo di errore ed alla intensità del vento al cubo.