

Notebook Assignment Rural – Berveglieri Filippo

Una volta ricevuta la traccia dell'assignment ho cercato di prendere confidenza con il file .gpkg prima installando il software QGIS e poi convertendo il file in un file .geojson.

Avere un file .geojson, molto simile al formato .json, mi consentiva una maggior compatibilità con il linguaggio Python che poi ho utilizzato per svolgere le operazioni di pulizia del dataset e di interpolazione.

Struttura del codice:

Innanzitutto, ho suddiviso il codice in diversi metodi per alcuni semplici motivi: mi consentiva di gestire meglio la complessità del sistema, per riutilizzare eventualmente alcune parti di codice e per poter testare il codice più facilmente.

esplorazione_dati.py

Questo metodo contiene la definizione della funzione utile a stampare il dataframe che gli viene passato. All'interno di questa funzione posso definire i dettagli di visualizzazione del dataframe come, nel mio caso, la richiesta di visualizzare tutte le righe e le colonne

main.py

Nel main, che è il core del mio progetto, inizialmente importo tutte le funzioni dichiarate in altri metodi in modo da poterle utilizzare. Successivamente, dentro alla funzione, queste funzioni "esterne" vengono chiamate. Dentro *punti_interpolazione* dovrebbero essere contenute le coppie rappresentanti i punti da interpolare attraverso i risultati dei valori noti ma, non avendo capito dove potessi trovare i punti da interpolare ho messo 4 punti casuali nelle vicinanze dei terreni contenenti i punti noti. L'interpolazione l'ho gestita con un blocco try-catch che restituisce i punti interpolati oppure, in caso di errore, il messaggio dell'errore. Infine, nel main, sono andato a stampare la matrice di correlazione: uno strumento che serve ad indicare se due colonne sono correlate tra loro e in che modo lo sono. La matrice di correlazione è utile per predire i punti sui valori da interpolare.

pulizia_dati.py

In *pulizia_dati* ho creato le funzioni per effettuare, appunto, la pulizia del dataset. La prima funzione chiamata è quella che andava a separare la colonna "data/ora" in due colonne distinte "data" e "ora". Dopo aver fatto ciò ho trasformato il separatore da virgola a punto nelle colonne "Longitudin" e "Latitudine" in modo da poter trasformare la colonna in un tipo numerico, utile sia per la matrice di correlazione che per l'interpolazione. Dopodiché nella funzione *datetime* ho cambiato la lingua in cui erano scritti i mesi, ho unito le due colonne e le ho trasformate in una colonna di tipo *datetime*. L'ultima funzione riguarda il processo di eliminazione dei valori NaN presenti nella colonna umidità che possono alterare le analisi.

interpolazione_idw_wrapper.py

Qua specifico che l'interpolazione ho scelto di farla con il metodo dell'Inverse Distance Weighting (IDW). Ho scelto di utilizzare questo metodo perché, non avendo mai fatto interpolazione spaziale, mi sembrava il più semplice da apprendere rispetto ad altri algoritmi come il TIN o il diagramma di Voronoi.

Per terminare l'assignment si sarebbe dovuto confrontare i risultati dell'interpolazione con un indice medio di vigore vegetazionale. Non avendo, però, i dati da interpolare non posso chiaramente fare questo passaggio. In ogni caso avrei dovuto valutare se i valori calcolati con l'algoritmo IDW sono correlati in qualche modo con l'indice medio di vigore vegetazionale.

Per eseguire il mio software è sufficiente scaricare il file .zip con i sorgenti, estrarlo, aprire i singoli file con un qualsiasi IDE compatibile (personalmente per lo sviluppo ho utilizzato PyCharm) e fare "run" sulla funzione main.

Per la parte di testing io ho testato il codice utilizzando la tecnica dello "unit testing" che è quella che consiglio.

Conclusioni:

La traccia dell'assignment era una traccia molto interessante che andava a toccare tematiche reali e interessanti e si è rivelata, specie nella prima parte, molto aderente alla parte di Python che conoscevo. Gli algoritmi di interpolazione spaziale non li conoscevo e per questo ho utilizzato l'algoritmo più immediato data la scarsità di giorni a disposizione e mi sarebbe piaciuto testarli con coordinate reali per portare a termine l'assignment. La parte inerente all'indice medio di vigore vegetazionale non l'ho potuta approfondire non avendo i punti interpolati anche se ho intuito fosse un modo per valutare la bontà dell'algoritmo,