

Master Universitario in Analisi Dati per la Business Intelligence e Data Science

Titolo tesi: Elaborazione di modelli previsionali relativi alle colture agricole attraverso l'analisi delle serie storiche, dei dati agrometeorologici e satellitari

Autore: Fracassi Federico

Anno Accademico: 2020/2021

## **Abstract**

Dal momento che le aziende agricole sono chiamate a operare in un ambiente economico competitivo e caratterizzato da una forte turbolenza, occorre adottare politiche gestionali dinamiche, capaci di fornire reazioni tempestive di fronte ai continui cambiamenti a livello tecnologico, organizzativo e ambientale. D'altra parte, la crescente attenzione all'ambiente e alla qualità delle produzioni rende rilevante l'adozione di tecniche integrate di difesa delle colture agricole. I modelli previsionali, capaci di allertare l'utenza agricola in caso di specifiche condizioni atmosferiche e climatiche, offrono di fatto la possibilità di orientare le scelte agronomiche verso determinate soluzioni colturali.

Traendo spunto dall'articolata letteratura scientifica in tema di applicazioni del Data Mining (DM) per la gestione integrata dell'informazione ambientale e territoriale, nel presente lavoro si fa riferimento alla valorizzazione dei dati puramente spaziali e spazio-temporali, al fine di favorire l'adozione di pratiche gestionali ed elaborare un modello previsionale per le rese delle colture presenti nel territorio piemontese sulla base dell'analisi delle serie storiche. In particolare, partendo dalle problematiche legate all'agricoltura di precisione, si focalizza l'attenzione sulla classificazione delle superfici coltivate attraverso vari algoritmi di DM, illustrandone le potenzialità utili alla definizione di processi decisionali in fase di gestione e pianificazione di interventi.

Dopo aver individuato e scaricato i dati necessari per l'elaborazione dei dati, sono state effettuate le seguenti operazioni:

- Data cleaning: utilizzando una serie di funzionalità presenti su Excel, sono rinominate le etichette di colonna in modo che fossero compatibili con i software utilizzati, divise le celle unite e ricodificati i dati riferiti alla stessa tipologia di coltura ma classificati diversamente nel corso degli anni
- Analisi Distribuzione Colture e Modelli Agrometeorologici attraverso:
  - Una serie di procedure presenti su SAS (PROC FREQ, PROC MEANS, PROC SGPLOT, PROC CORR e PROC REG), sono stati stimati modelli di regressione lineare multipla, gestendo opportunamente i valori missing
  - Una serie di algoritmi di classificazione presenti su KNIME (K Nearest Neighbor, Decision Tree, Random Forest), è stata comparata l'accuratezza, attraverso le matrici di confusione che mostrano la corretta classificazione, per stimare il miglior modello di previsione possibile