



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO

**dib** DIPARTIMENTO DI  
INFORMATICA

---

**Dottorato di ricerca in Informatica e Matematica  
XXXII ciclo**

**Progetto di ricerca**

**Dottorando:** Dott.ssa Alessandra Legretto

**Tutor:** Prof. Paolo Buono

Firma del dottorando \_\_\_\_\_

Firma del tutor \_\_\_\_\_

## **1. Titolo della ricerca:**

Modelli, tecniche e strumenti per l'analisi predittiva di grandi quantità di dati, attraverso tecniche visuali interattive.

## **2. Area nella quale si inquadra la ricerca:**

Predictive Visual Analytics e Big Data

## **3. Motivazioni e obiettivi della ricerca**

I ricercatori che operano nell'ambito dell'Information Visualization, insieme a chi si occupa di Data Mining, stanno dando una forte spinta a una disciplina emergente che è la Visual Analytics. Nonostante ad oggi sia stata studiata un'ampia varietà di tecniche visuali interattive, esse sono utilizzate soprattutto per la presentazione di risultati di algoritmi di Data Mining.

La novità nelle tecniche di Visual Analytics che saranno studiate ed implementate durante la ricerca nei tre anni di dottorato consiste nell'uso di visualizzazioni interattive per fornire all'analista ulteriore conoscenza sui dati e per consentirgli di guidare il processo di analisi.

L'interesse della ricerca nell'analisi di dati oggi è rivolto alla scelta della migliore tecnica che abiliti il data scientist all'attività di analisi. Le persone oggi sono sempre più connesse, gli analisti devono spesso affrontare problemi lavorativi mentre sono in mobilità, pertanto è di notevole interesse studiare come tali tecniche, di tradizionale appannaggio dei desktop, che utilizzano anche video di grandi dimensioni, possano essere adattate ai piccoli schermi degli smartphone.

Una innovazione è veramente tale quando migliora la vita delle persone. Obiettivo è implementare tecnologie innovative ed efficienti, utilizzabili dagli utenti finali per migliorare la vita del cittadino ed il business per le imprese. L'uso degli strumenti visuali come supporto al data scientist per prendere decisioni rappresenta un modo per migliorare la sua attività lavorativa. Si potranno verificare i vantaggi che l'introduzione di tecniche di Visual Analytics portano nel lavoro quotidiano dei data scientist prendendo in considerazione i risultati prodotti in un dato intervallo di tempo e confrontandoli con quelli prodotti da altre tecniche.

## **4. Stato dell'arte**

La ricerca intende investigare modelli, tecniche e strumenti di Predictive Visual Analytics, che hanno grosse potenzialità di risultare molto promettenti per supportare l'analista di Big Data.

Con il termine Big Data si intendono raccolte di dati di notevole ampiezza per cardinalità e/o per dimensione, per i quali i tradizionali metodi di analisi sono inadatti. Visual Analytics è una disciplina recente che ha l'obiettivo di esplorare le migliori sinergie tra metodi di Data Mining (basati su algoritmi statistici o di Machine Learning) e tecniche di visualizzazione interattive, al fine di identificare pattern significativi nei dati e consentire all'analista una migliore comprensione sia dei dati che dei processi complessi basati su di essi, consentendogli di prendere decisioni più consapevoli.

L'analisi predittiva è una parte importante dell'analisi dei dati. Statistica e Machine Learning forniscono vari metodi predittivi principalmente per scopi di regressione e classificazione. Recenti ricerche hanno iniziato a considerare approcci visuali interattivi allo scopo di migliorare l'analisi

predittiva. Tecniche di Visual Analytics usano visualizzazioni interattive per fornire all'analista ulteriore conoscenza sui dati e per consentirgli di guidare il processo di analisi. In quest'ambito ci sono vari problemi aperti sui quali si intende investigare, ad esempio:

- come una visualizzazione supporta il data scientist nella comprensione dei modelli predittivi?
- come assicurare agli utenti finali che i modelli predittivi si basano su caratteristiche rilevanti per il dominio di interesse?
- quali sono i processi più adatti per assicurare l'utilità dei modelli predittivi?
- come le visualizzazioni possono aumentare la fiducia dei data scientist nei risultati ottenuti?

## **5. Approccio al problema**

Come prima fase di approccio al problema si prevede uno studio dello stato dell'arte e della letteratura nell'ambito della Predictive Visual Analytics e dei Big Data. Successivamente, in ragione del posizionamento sul mercato della tematica dell'analisi di Big Data e dell'importanza che essa ricopre per l'azienda, l'attività formativa e di ricerca intende sviluppare abilità di analisi di dati e di arricchire tali abilità mediante lo studio e lo sviluppo di tecniche visuali interattive (Visual Analytics), in modo da consentire all'analista di acquisire ulteriore conoscenza sui dati e dargli la possibilità di guidare il processo di analisi. Ci si attende, inoltre, che come dottoranda, possa all'interno dell'impresa le conoscenze e buone pratiche acquisite, colmando una lacuna di competenze attualmente evidenziata a livello aziendale.

Tale trasferimento di competenze assume importanza strategica nell'ottica degli investimenti aziendali a medio e lungo termine ed in considerazione delle progettualità in cantiere sia in termini di iniziative di ricerca e sviluppo, che di commesse nei settori di mercato di punta. A tal proposito ed a titolo di esempio, in tema analisi di Big Data si annoverano tra gli interessi aziendali gli scenari del monitoraggio e della gestione del rischio del credito in ambito finanziario, e la determinazione di indicatori funzionali al supporto alle decisioni nel contesto della valutazione di dati relativi e consumi energetici, livelli dell'inquinamento e di sicurezze delle "smart cities".

Con il Prof. Bertini è stato definito che, durante il periodo di ricerca di sei mesi nel laboratorio InfoVis, si lavorerà principalmente su Predictive Visual Analytics. In particolare, durante tale periodo, collaborando con i ricercatori della NYU, verrà proseguita l'attività di progettazione, implementazione e valutazione User-Centred di prototipi che utilizzano tecniche di Predictive Visual Analytics. Inoltre, verranno acquisite specifiche abilità sui metodi di valutazione di tecniche di visualizzazione che consentiranno di stabilire se tali tecniche sono realmente efficaci e efficienti, e dunque utilizzabili, nei contesti reali che sono quelli significativi per le aziende. È, infatti, ben noto che le aziende, piuttosto che alle ricerche per sé, sono maggiormente interessate alla validità applicativa di quanto si sviluppa in accademia, per poterlo utilizzare nei loro contesti lavorativi e di mercato ed attenerne un vantaggio in termini economici e di competitività.

## **6. Risultati attesi**

La ricerca è incentrata su modelli, tecniche e strumenti per l'analisi predittiva di grandi quantità di dati (Big Data) attraverso tecniche visuali interattive. La proposta di ricerca è coerente con la Strategia Nazionale di Specializzazione Intelligente, in particolare con l'area tematica "Agenda Digitale, Smart Communities, Sistemi di mobilità intelligente". Uno degli obiettivi di tale area tematica è implementare tecnologie innovative ed efficienti, adatte agli utenti finali, integrando informazioni e generando intelligenza, in modo da migliorare la vita del cittadino ed il business per le imprese. Si sottolinea che Big Data Analytics è una delle 10 tecnologie abilitanti previste per il piano nazionale Industria 4.0, con l'obiettivo di ottimizzare prodotti e processi produttivi.

L'azienda Links Management & Technology (per brevità denominata Links) è particolarmente interessata a tale ricerca perché avverte una notevole esigenza di acquisire conoscenze sia sui metodi di analisi di Big Data che sulle tecniche di visualizzazione a supporto dell'analisi di Big Data. Le possibili applicazioni nell'ambito dei progetti che sviluppa sono varie. Links intende applicare i risultati della ricerca negli ambiti finanziario e delle Smart Cities, contesti di mercato rispetto ai quali l'azienda ha maturato una pluriennale esperienza e in cui intende continuare ad investire a medio e lungo termine.

Il tema dell'analisi di Big Data è estremamente attuale ed assume notevole rilevanza per il business, non solo con un obiettivo di mero adeguamento ai trend di innovazione ma anche per una concreta esigenza in termini di competitività, come testimoniano i molteplici ambiti di mercato nei quali trova applicazione: da quello finanziario al sanitario, dalla logistica ai trasporti, alla produzione industriale, dove è necessario che le aziende diano valore al patrimonio informativo esistente, in modo da riuscire a competere meglio in un mercato globale.

In tale contesto, a fronte di un grande sforzo profuso nella ricerca di sempre nuove ed eterogenee fonti informative, in termini di volume, varietà e velocità di aggiornamento delle informazioni (tre "V" dei Big Data [1]); da mettere a disposizione anche in modalità "open", è necessario un altrettanto significativo impegno nello sviluppo di una capacità analitica e di competenze in termini di metodologie, tecniche e di tecnologie per l'analisi predittiva, la simulazione di scenari, l'interpretazione delle informazioni per il supporto alle decisioni e la presentazione dei dati. Il tema della ricerca consente, da un lato, di generare competenze finalizzate all'estrazione di valore (quarta "V" dei Big Data) dai dati grazie ad analisi finalizzate alle previsioni ed al supporto alle decisioni, dall'altro di offrire un approccio innovativo all'accesso ai dati, basato su tecniche di visualizzazione che tengano conto, tra gli altri, di aspetti quali i canali ed i dispositivi di erogazione e fruizione delle informazioni.

Relativamente alla situazione illustrata, è evidente che l'industria è fortemente interessata alle competenze che il dottorando acquisirà e che saranno, pertanto, immediatamente spendibili nel mondo del lavoro.

## **7. Fasi del progetto**

L'attività di ricerca è organizzata secondo le seguenti fasi principali.

Studio dei metodi predittivi tradizionali basati su tecniche statistiche e algoritmi di Machine Learning.

Analisi dello stato dell'arte, relativamente a metodologie, tecniche e strumenti esistenti in letteratura e nell'ambito commerciale per l'analisi visuale di dati (Visual Analytics).

Identificazione dei compiti significativi per un analista al fine di individuare le tecniche più indicate per supportare tali compiti, contestualizzandoli nei domini di interesse.

Contestualizzazione dell'attività di ricerca nell'ambito di cantieri progettuali di interesse di mercato per Links, interagendo con i ricercatori di Links e studiando documenti, dati e software necessari alla comprensione del dominio applicativo ed allo svolgimento dell'attività di ricerca in accordo con le esigenze aziendali.

Progettazione e implementazione e valutazione di tecniche innovative di Predictive Visual Analytics, anche sulla base dell'attività che il dottorando svolgerà presso il centro di ricerca estero, dove si recherà per un periodo di almeno sei mesi.

I anno

- studio dello stato dell'arte relativo a Big Data, caratteristiche e approcci all'analisi;

- studio dei metodi predittivi tradizionali basati su tecniche statistiche e algoritmi di Machine Learning;
- studio dello stato dell'arte su metodologie, tecniche, strumenti esistenti in letteratura e nell'ambito commerciale per l'analisi visuale di dati (Visual Analytics);

#### II anno

- identificazione dei compiti significativi per un analista al fine di individuare i requisiti per sviluppare le tecniche di VA più adatte per supportare tali compiti, contestualizzandoli nei domini di interesse;
- progettazione, implementazione e valutazione, possibilmente coinvolgendo un campione di utenti finali, dei primi prototipi che utilizzano le tecniche proposte; tutto ciò è in linea con una metodologia di progettazione User-Centred;
- proseguo dell'attività iterativa di progettazione, implementazione e valutazione di prototipi durante il periodo di ricerca presso il laboratorio estero;

#### III anno

- adattamento dei prototipi sviluppati ai domini di interesse dell'azienda;
- definizione di modelli/principi/linee guida per creare strumenti efficaci ed efficienti di Predictive Visual Analytics;
- sperimentazione delle tecniche sviluppate attraverso test con utenti finali.

## 8. Valutazione dei risultati

La valutazione dei risultati avviene principalmente seguendo due vie: una prevede il coinvolgimento degli utenti finali per la valutazione dell'esperienza d'uso delle tecniche di visualizzazione, l'altra utilizza metriche di valutazione dell'accuratezza predittiva.

L'utilità di una tecnica deve essere valutata con il coinvolgimento degli utenti. Saranno effettuati ed confronti con altri approcci in condizioni controllate ed esperimenti controllati, per testare vari parametri della tecnica di visualizzazione.

La valutazione dell'accuratezza e della bontà del modello predittivo sarà effettuata identificando adeguate metriche che consentono di valutare veri positivi e i veri negativi (istanze correttamente classificate dal modello) ma anche i falsi positivi e i falsi negativi (istanze non correttamente classificate dal modello) oltre che con l'uso di precision e recall, combinate con metriche più complesse.

## 9. Eventuali referenti esterni al Dipartimento

Dott.ssa Palmalisa Marra supervisore Links Management and Technology.

Dott. Enrico Bertini, supervisore presso NY Univesity. Ricercatore molto attivo nell'ambito della visualizzazione di informazioni e Visual Analytics, interessato particolarmente allo studio e all'implementazione di tecniche e strumenti per l'analisi esplorativa di dati, che tiene conto delle caratteristiche della percezione umana per produrre visualizzazioni interattive efficaci.

## 10. Riferimenti bibliografici

1. Keim, Andrienko, Fekete, et al. "Visual analytics: Definition, process, and challenges". *Lecture notes in computer science*, 4950 (2008), 154-176.
2. Keim, Mansmann, Thomas. "Visual Analytics: How Much Visualization and How Much Analytics?". *ACM SIGKDD Explorations Newsletter* 11.2 (2010): 5-8.
3. Lu, Junhua, et al. "Recent progress and trends in predictive visual analytics." *Frontiers of Computer Science* (2017): 1-16.
4. E. D. Keim, J. Kohlhammer, and G. Ellis. *Mastering the information age: Solving problems with visual analytics*, eurographics association, 2010.
5. J J.J. Thomas and K.A. Cook, eds., *Illuminating the Path: The Research and Development Agenda for Visual Analytics* , IEEE CS Press, 2005; <http://nvac.pnl.gov/agenda.stm>.