



UNIVERSITA' POLITECNICA DELLE MARCHE
FACOLTA' DI INGEGNERIA

Corso di Laurea Triennale in Ingegneria Gestionale

**ANALISI E GESTIONE DELLA QUALITÀ NEL SETTORE
MANIFATTURIERO**
**QUALITY ANALYSIS AND MANAGEMENT IN THE MANUFACTURING
SECTOR**

Relatore: Chiar.mo/a
Prof. **Maurizio Bevilacqua**

Tesi di Laurea
Toro Giorgia

Correlatore: Chiar.mo/a
Prof. **Sara Antomarioni**

A.A. 2021 /2022

INDICE

Introduzione.....	5
-------------------	---

CAPITOLO 1: CONTROLLO QUALITÀ E PROGRAMMAZIONE DELLA PRODUZIONE

1.1. La qualità.....	7
1.2. Evoluzione del concetto di qualità.....	7
1.3. Norme ISO 9000.....	9
1.4. Qualità e industrie.....	10
1.4.1. Analisi dei dati.....	11
1.4.2. Controllo qualità.....	12
1.5. Programmazione della produzione.....	13

CAPITOLO 2: L'AZIENDA MATRIX E IL PROCESSO DI LAVORAZIONE DELLA CERAMICA

2.1. L'azienda.....	14
2.1.1. Progetti futuri.....	14
2.2. Il processo di manifattura della ceramica.....	15
2.2.1. Preparazione dell'impasto.....	15
2.2.2. Foggiatura.....	16
2.2.3. Trattamento delle superfici.....	17
2.2.4. Fase di cottura.....	17

CAPITOLO 3: REGOLE DI ASSOCIAZIONE E RAPIDMINER

3.1. Data Mining.....	19
3.1.1. Regole di associazione.....	21
3.1.2. L'Algoritmo Apriori.....	22
3.1.3. L'Algoritmo FP-Growth.....	23
3.2. Rapidminer.....	24
3.2.1. Processi realizzati.....	26

CAPITOLO 4: DATASET E ANALISI DEI RISULTATI

4.1. Dataset iniziale.....	29
4.2. Estrazione delle regole e considerazioni.....	30
4.2.1. Regole associative primo processo.....	30

4.2.2. Regole associative secondo processo.....	32
CONCLUSIONI.....	36
BIBLIOGRAFIA.....	37
SITOGRAFIA.....	38
RINGRAZIAMENTI.....	39

INTRODUZIONE

Il mercato attuale è caratterizzato da clienti sempre più attenti ed esigenti, che richiedono alle aziende di sostenere la loro capacità competitiva con prodotti sempre più qualificati. In altri termini, fattori come la Qualità e la prontezza di risposta sono fattori determinanti per l'intero processo aziendale e per la competitività sul mercato. L'approccio alla gestione della qualità all'interno di piccole e grandi imprese si è evoluto nel tempo passando da una semplice ispezione visiva a metodi sempre più moderni e tecnologici che vedono l'utilizzo di software informatici. La presente tesi ha lo scopo di effettuare una dettagliata analisi della qualità di prodotti sanitari in ceramica realizzati dall'azienda Matrix srl, con la successiva analisi sulla pianificazione della produzione futura.

L'intero progetto di analisi è stato supportato da un software chiamato Rapidminer che ha permesso l'estrazione di regole associative interessanti in tempi molto brevi. L'intero elaborato è strutturato in quattro capitoli.

Nel primo capitolo è presente una panoramica del concetto di Qualità, sottolineando la sua evoluzione storica a partire dal periodo antecedente la Rivoluzione Industriale fino ai tempi attuali. In particolare, sono stati fatti riferimenti ai vari studiosi che nel tempo hanno elaborato il concetto di Qualità Totale, da cui nascono le Norme ISO 9000. Inoltre, è stata posta particolare attenzione sulle moderne tecniche di gestione della qualità che, a partire dall'analisi dei dati in forma grezza, costruiscono modelli predittivi in grado di scoprire delle ricorrenze nei sistemi ed effettuare previsioni sui futuri cicli produttivi.

Il secondo capitolo dell'elaborato, descrive l'azienda Matrix Modellerie con cui è stato effettuato il tirocinio e l'intero lavoro di analisi, ma descrive anche il processo di lavorazione della ceramica da cui nascono i vari prodotti sanitari che l'azienda offre sul mercato.

Il terzo capitolo analizza il progetto svolto in accordo con l'azienda. Parte dalla descrizione del concetto di Data Mining, con particolare riferimento agli algoritmi che permettono l'estrazione di regole associative (Apriori e FP-Growth). Come già accennato, l'estrazione di regole associative costituisce il fulcro dell'intero progetto di tirocinio. Inoltre, in questo capitolo, è presente una panoramica del software Rapidminer con la descrizione di tutte le funzionalità utilizzate per la costruzione dei processi che hanno permesso l'estrazione delle regole.

L'ultimo capitolo della tesi ha l'obiettivo di presentare i risultati e le analisi effettuate una volta processati i dati forniti dall'azienda. In particolare, si pone l'attenzione sulle funzionalità delle regole associative estratte che hanno permesso di formulare previsioni sulla pianificazione della produzione futura.

CAPITOLO 1: controllo qualità e programmazione della produzione

1.1: La qualità

Il concetto di qualità, al giorno d'oggi, è ampiamente utilizzato in qualsiasi ambito. Si parla di “qualità della vita” per indicare genericamente il benessere di un individuo, o anche si parla di “qualità di un prodotto” per indicare il grado di soddisfazione che il prodotto fornisce al consumatore finale (G. Bizzarri, C. Guarnieri).

È quindi molto complesso dare una definizione unica e specifica del concetto di qualità.

Questo elaborato tratterà unicamente i temi dedicati alla gestione della qualità in ambito aziendale.

Al giorno d'oggi il concetto di qualità è diventato uno dei fattori determinanti nel processo decisionale dei consumatori, soprattutto quando si tratta di scegliere prodotti o servizi in competizione tra loro. Di conseguenza, il fattore qualità è un elemento fondamentale nel determinare il posizionamento sul mercato delle aziende ed è ormai uno dei punti cardine dell'intero processo di pianificazione.

1.2: Evoluzione del concetto di qualità

Al fine di valutare l'importanza che la qualità ha assunto nel tempo, è bene presentare la sua evoluzione storica.

Prima della Rivoluzione Industriale, la qualità di prodotti o servizi era un problema che riguardava soltanto le classi più abbienti in quanto un'elevata qualità si traduce in prezzi maggiori. Con l'avvento della Rivoluzione Industriale questa filosofia non subì eccezionali cambiamenti; in un primo momento la meccanizzazione portò ad un aumento della quantità prodotta e, parallelamente, ad un aumento della qualità.

Tuttavia, questo processo subì un notevole rallentamento con l'avvento della produzione in serie che aveva come obiettivo la massimizzazione dei volumi di produzione, mettendo in secondo piano la qualità dei prodotti.

Ben presto la produzione in serie cominciò ad evidenziare un problema fondamentale: l'enorme quantità di pezzi difettosi e di conseguenza il loro elevato costo. Infatti, un operatore assegnato alla realizzazione di una singola operazione non aveva una visione generale della produzione completa e di conseguenza non era in grado di valutare l'effettiva qualità del prodotto finale.

Per far fronte a questa difficoltà, intorno agli anni '30, venne introdotto il “controllo statistico della qualità” da parte di Walter A. Shewhart. Per controllo statistico della qualità si intende esclusivamente la valutazione della conformità all'uso del prodotto finale che viene effettuata utilizzando il metodo del campionamento statistico, senza però effettuare preliminari analisi di pianificazione o controllo (F. Marmura, 2012).

Solo dopo la Seconda Guerra mondiale, si manifestò nelle aziende la necessità di pianificare studi sulla qualità, in quanto si diffuse la consapevolezza che i costi sulle ricerche finalizzate alla qualità fossero sicuramente inferiori ai costi relativi la produzione di pezzi difettosi che non potevano essere immessi sul mercato.

Negli anni '50 alcuni studiosi, considerati i “padri della qualità”, come A.V Feigenbaum, J.M Juran e E.V Deming introdussero il concetto di qualità totale inteso come soddisfazione del cliente e miglioramento continuo dei processi.

In particolare, il matematico statunitense Feigenbaum propose un metodo sulla gestione della qualità, chiamato “Total Quality Control”, che ha come obiettivo il miglioramento continuo dell'intera struttura aziendale. Tale strategia è basata sul presupposto che il successo e la sopravvivenza di un'azienda sia strettamente legata al soddisfacimento del consumatore finale.

Poiché i consumatori sono alla continua ricerca di un livello di qualità sempre più alto dei prodotti, è necessario che le aziende si pongano come obiettivo proprio il miglioramento continuo della qualità. Il Total Quality Control rappresenta uno strumento di gestione di qualità caratterizzato dalle attività di analisi, controllo e miglioramento del processo. (F. Marmura, 2012). (A.V. Feigenbaum, 1961).

Secondo Joseph M. Juran (1904-2008), ingegnere e consulente gestionale, bisogna concentrarsi principalmente su due aspetti fondamentali:

- il primo riguarda le caratteristiche dei prodotti: un'alta qualità si traduce in un elevato numero di caratteristiche che soddisfano le esigenze dei consumatori;
- il secondo punto intende la qualità come adeguatezza all'uso e come conformità alle caratteristiche (zero difetti del prodotto).

Juran pone l'attenzione sul mercato, proponendo di realizzare un prodotto “su misura” considerando le volontà e le aspettative dell'utente finale, aggiungendo però come altra componente di analisi la gestione dei costi.

La Trilogia della Qualità di Juran è un approccio alla gestione che si basa su tre principi:

- pianificazione: processo relativo lo sviluppo vero e proprio dei prodotti/servizi;

- controllo: fase in cui vengono effettuate tutte le procedure di collaudo e di controllo sulle performance e sulla qualità dei prodotti/servizi realizzati;
- miglioramento: fase in cui, una volta immesso il prodotto sul mercato, vengono effettuati studi di ricerca e sviluppo con l'obiettivo di realizzare progetti di miglioramento. (J. Juran,1997)

William E. Deming (1900-1993) utilizza un approccio prettamente statistico per definire il concetto di qualità. Grazie a Deming, la definizione di qualità si estende all'affidabilità e alla durabilità dei prodotti.

Secondo Deming un qualsiasi processo produttivo può essere visto come un ciclo che prevede quattro fasi fondamentali con l'obiettivo di raggiungere il massimo livello della qualità: *plan* (progettare, pianificare), *do* (agire, realizzare), *check* (controllare), *act* (regolare).

- *Plan*: fase in cui si definiscono gli obiettivi e si pianificano i miglioramenti da apportare ai processi;
- *Do*: realizzazione dei prodotti/ servizi;
- *Check*: fase di controllo e collaudo;
- *Act*: fase di consolidamento e standardizzazione dei risultati;



Figura 1: Illustrazione del ciclo di Deming

Il ciclo di Deming costituisce ancora oggi uno dei fondamenti della pianificazione ed il punto di partenza per tutti gli studi effettuati negli ultimi anni sulla qualità e sulla pianificazione della produzione. Con Deming inizia infatti un processo di continua evoluzione del concetto di qualità fino ad arrivare ad una definizione che unisce i diversi punti di vista. (E. Deming, 2012), (P. Senni).

Si sviluppa così il concetto di Sistema di Gestione della Qualità che verrà codificato nelle norme ISO serie 9000.

1.3: Norme ISO 9000

Il Sistema di Gestione della Qualità, realizzato secondo i principi delle norme UNI EN ISO 9000, rappresenta un insieme di elementi che hanno l'obiettivo di mantenere sotto controllo un'azienda mediante la definizione di precise responsabilità, regole comportamentali e

metodi operativi e di controllo, con lo scopo di assicurare che quanto realizzato sia in linea con le politiche e gli obiettivi aziendali. (D. Grisot,2006).

La capacità di un'azienda di gestire la qualità non è valutata indirettamente, attraverso la realizzazione del prodotto finale, ma direttamente, cioè attraverso l'analisi dell'efficienza dell'intero sistema organizzativo. La novità introdotta da queste norme è costituita da un particolare approccio ai processi, che sposta il centro dell'attività aziendale dal solo sistema produttivo, al sistema gestionale nella sua interezza comprendendo anche il fattore umano. Le norme non forniscono dei requisiti per lo sviluppo di un Sistema di Gestione della Qualità (non dicono come deve essere fatto un sistema organizzativo), ma forniscono una serie di requisiti e suggerimenti su cosa deve essere fatto per poter sviluppare un efficace sistema di gestione della qualità. Bisogna, quindi, prevenire gli errori piuttosto che correggerli e per farlo è necessario controllare costantemente l'intero processo prestando particolare attenzione a tutte quelle variabili che hanno maggiori criticità.

Le norme ISO 9000 sono universali e la loro applicabilità non dipende dalle dimensioni o dal settore in cui opera l'azienda, in quanto i principi che le regolano non sono applicabili ai prodotti veri e propri ma alle aziende che li producono. (D. Grisot, 2006)

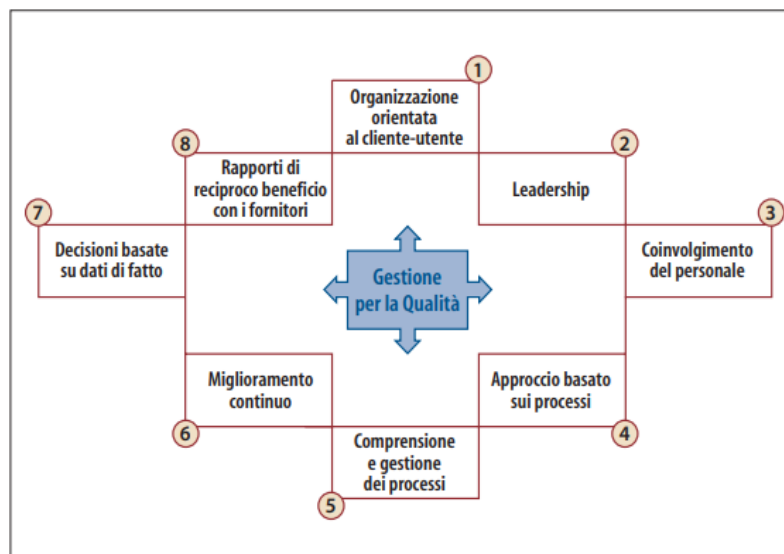


Figura 2: gli 8 principi della qualità (G. Bizzarri, C. Guarnieri)

1.4: Qualità e industrie

Il monitoraggio e la valutazione della qualità della produzione è sempre stato uno dei principali problemi delle aziende che hanno come obiettivo quello di riuscire a soddisfare le aspettative dei clienti e di rimanere competitive sul mercato. Prima dell'avvento dell'*Industria 4.0*, il controllo qualità nelle aziende era affidato a operatori che, tramite ispezione visiva, erano chiamati ad individuare eventuali pezzi difettosi (non conformi a determinate caratteristiche).

Con il termine *Industria 4.0* si intende la tendenza, da parte delle imprese, di affiancare al fattore umano tutta una serie di nuove tecnologie produttive automatizzate per migliorare le condizioni di lavoro e aumentare la produttività degli impianti e la loro qualità. (M. Bevilacqua, 2017/2018).

Per quanto riguarda il miglioramento della qualità produttiva, ancora oggi vi è una crescente domanda, tra le imprese, per l'adozione di sistemi automatici per il rilevamento dei difetti. Tuttavia, il problema principale quando si utilizzano questi sistemi è lo squilibrio dei *data-set* industriali: spesso i campioni privi di difetti (normali) superano di gran lunga i campioni difettosi (anormali) grazie ai continui miglioramenti adottati dalle industrie manifatturiere. Di conseguenza si è pensato di utilizzare modelli predittivi, come i modelli di *Deep Learning* con strutture basate su reti neurali, per la previsione della qualità. La costruzione di un modello è un processo che permette di passare dalla semplice raccolta e organizzazione dei dati osservati, ad analisi, interpretazione e previsione del comportamento futuro del sistema. (L. Screpanti, 2022).

1.4.1: Analisi dei dati

Il primo problema da affrontare, durante la costruzione di modelli predittivi su dati per la previsione della qualità, è la scelta delle variabili di input da fornire al modello per poterlo addestrare. Addestrare un modello (*fase di training*) significa utilizzare degli algoritmi di apprendimento automatico per trovare delle relazioni significative tra le variabili in esame. I principali metodi di selezione delle variabili di input si basano sulla conoscenza del processo e sull'esperienza degli operatori impiegati nel controllo qualità.

Per avere dei modelli ben strutturati bisogna partire dal concetto di causalità tra variabili di processo e variabili di qualità. (Lee Yao, Zhiqiange).

Si effettuano delle analisi di causalità per la determinazione delle variabili di input da fornire al modello. In primo luogo, l'analisi di causalità permette di migliorare l'interpretabilità del modello di previsione della qualità, intesa come il grado con cui ciascuna variabile influenza la qualità stessa. Successivamente, con un'analisi di questo tipo si avrà una migliore capacità di rilevamento di difetti ed un modello predittivo più robusto e stabile rispetto ai tradizionali metodi basati sulla correlazione.

I metodi di analisi di qualità si basano sui due algoritmi: la causalità di Granger (GC) e l'entropia di trasferimento (TE). Il primo viene calcolato applicando un modello autoregressivo (AR) ai dati del processo, ma è limitato a un processo lineare e stazionario.

Il secondo invece, da un lato prende in considerazione le variabili non lineari del processo, dall'altro però è un metodo scomodo per l'analisi dei dati a causa dei limiti del regime stazionario, l'alta sensibilità del processo e l'elevata complessità computazionale.

Di conseguenza, negli ultimi anni, l'algoritmo di causalità di Granger è stato migliorato per dedurre la relazione causale tra variabili di processo stazionarie e non lineari.

(M. Chaabi et al., IAES International Journal of Artificial Intelligence).

1.4.2: Controllo qualità

Come già detto, avere un efficace sistema di gestione e controllo qualità è per le aziende un punto fondamentale per la loro sopravvivenza. (D. Grisot, 2006).

In particolare, il controllo qualità è un'attività puramente ispettiva inserita all'interno del processo produttivo. Attraverso questa attività, le aziende sono in grado di valutare se i prodotti realizzati sono conformi alle specifiche definite dal modello predittivo (dedotto dall'analisi dei dati).

Alcuni pezzi vengono definiti "difettosi" quando non rispettano le specifiche definite e possono essere o scartati a priori, oppure reindirizzati ad una nuova fase del processo chiamata ripasso.

È possibile effettuare diverse tipologie di controllo qualità:

- a inizio ciclo: per evitare di utilizzare già dall'inizio dei materiali non conformi;
- controlli di tipo "orizzontale": tra i reparti, nelle fasi intermedie di lavorazione;
- controlli "a ritroso": sui prodotti finiti, nella fase di collaudo.

Come si può intuire, non è consigliabile inserire il solo controllo qualità all'inizio o alla fine del ciclo produttivo in quanto controllare soltanto le conformità dei materiali che entrano in produzione non garantisce un alto livello della qualità del prodotto finale, né tanto meno permette di bloccare eventuali pezzi difettosi. D'altra parte, quando si inserisce il controllo qualità alla fine del ciclo e si evidenziano delle non conformità sui prodotti finiti, si avrà un elevato costo per il loro smaltimento o per la loro rilavorazione.

Quella del controllo qualità è un'attività utile in tutte quelle aziende che hanno come obiettivo il miglioramento continuo, in termini di qualità.

Per concludere si può affermare che affinché questa attività possa essere implementata correttamente in un sistema aziendale è necessario:

- Definire le caratteristiche da controllare sulla base dei risultati forniti dal modello predittivo (dimensionali, superficiali, strutturali, etc....)

- Stabilire in quali fasi del processo produttivo inserire l'attività di controllo e su quali componenti/prodotti;
- Eseguire periodicamente i controlli e le analisi;
- Motivare tutta la catena aziendale che fa parte di questa attività (dal singolo operatore all'intero reparto)

Attraverso una corretta implementazione del Sistema di Controllo Qualità nessun prodotto difettoso raggiungerà mai l'utente finale.

1.5: Programmazione della produzione

Le attività di analisi dei dati e di controllo qualità possono essere studiate anche in chiave di pianificazione. In particolare, una volta costruito il modello predittivo del sistema, che mette in relazione le variabili del processo e le variabili sulla qualità, è possibile fare delle previsioni sui futuri cicli produttivi.

Quando all'interno di un modello predittivo è possibile stabilire e identificare delle ricorrenze, esso diventa un modello matematico che può essere applicato anche ad altri data-set. (L. Screpanti, 2022).

Questa caratteristica evidenzia la duplice funzionalità del Sistema di Gestione e controllo Qualità: se da un lato permette di tenere sotto controllo la produzione identificando la presenza di eventuali pezzi difettosi, dall'altro consente di programmare la produzione per futuri cicli. A partire dalle “regole associative” (concetto che verrà ripreso in seguito) formulate dal modello predittivo ed effettuando un'attenta analisi di queste ultime, si è in grado di migliorare continuamente l'intero processo produttivo, obiettivo ultimo di ogni impresa manifatturiera (e non).

Il fine di questo elaborato è quello di applicare tutti i principi del controllo qualità sopra elencati in un progetto che vede protagonista un'azienda della provincia di Teramo, Matrix Srl, specializzata nella modellazione della ceramica. Di seguito è riportata una panoramica dell'azienda presa in esame e una descrizione del processo di lavorazione della ceramica.

CAPITOLO 2: L'azienda Matrix e il processo di lavorazione della ceramica

2.1: L'azienda

L'azienda Matrix Srl, situata in provincia di Teramo, nasce nel 2003 dall'unione di un team di esperti nel settore della modelleria ceramica sanitaria. Punto di forza dell'azienda è rappresentato dagli elevati standard qualitativi dei prodotti, che le hanno permesso di diventare partner tecnico dei principali marchi mondiali.

Matrix si pone come obiettivo la continua ricerca della migliore soluzione per il cliente, analizzando e prevedendo tutte le criticità inerenti ai processi produttivi e fornendo un supporto tecnico nel periodo post-vendita. Infatti, elemento centrale per l'azienda è rappresentato dal cliente e dalla sua soddisfazione.

I gruppi leader di mercato di grandi e piccole imprese su scala globale hanno trovato in Matrix un leale e valido partner tecnico e rinnovano anno dopo anno la fiducia in essa. (Matrixmodelleria.com)

2.1.1: Progetti futuri

Negli ultimi anni l'azienda Matrix si è posta come obiettivo di lungo periodo l'espansione dell'intero settore produttivo, a partire dall'ampliamento della manodopera. Inoltre, nel 2019, l'azienda ha dato il via ad un piano di investimenti che prevedono un sensibile aumento della superficie produttiva, l'implementazione di soluzioni tecnologiche da affiancare ai processi produttivi e la realizzazione di un'area museo che racconta la storia della ceramica dagli anni '60 ad oggi.

Tra i progetti di maggior successo dell'azienda Matrix vi è senza dubbio il progetto Etò ceramiche. (etoceramiche.it).

L'obiettivo del progetto è quello di unire l'artigianato classico con il contemporaneo design, per creare collezioni di sanitari all'avanguardia e di qualità.

Etò è diventato un brand vero e proprio, in cui lavorano artigiani esperti e giovani ragazzi volenterosi di imparare l'arte, che esporta le sue collezioni in tutto il mondo.

2.2: Il processo di manifattura della ceramica

Il processo di manifattura della ceramica comprende diverse fasi fondamentali:

- L'impasto: estrazione e lavorazione delle materie prime e lavorazione dell'impasto;
- Foggatura: per dare una forma generale;
- Trattamento delle superfici: rifiniture;
- Cottura.

(S.T Levi, 2010).

2.2.1: Preparazione dell'impasto

La materia prima alla base dell'intero processo di lavorazione della ceramica è l'argilla, roccia sedimentaria costituita principalmente da materiali argillosi. L'argilla è un materiale facilmente reperibile in quanto ampiamente diffuso nel territorio, che presenta diverse caratteristiche come la plasticità che ne consente la modellazione.

La materia prima della ceramica è formata da una miscela di componenti che possiamo suddividere in due classi in base al grado di granulometria di cui sono costituiti: si parla di *matrice* per indicare la parte più fine e di *clasti* per indicare quella più grossolana.

La fase di preparazione della ceramica comprende diverse operazioni: prima tra tutti è la fase di stagionatura dell'argilla, che dura circa un anno, ed ha l'obiettivo di aumentare le proprietà plastiche del materiale. Successivamente l'argilla stagionata verrà trasferita all'interno di un grosso turbo dissolutore per la fase di Scioglitura in cui sarà effettuato il lavaggio dell'argilla insieme all'acqua per eliminare i sali solubili al suo interno.

A seguito della Scioglitura, l'argilla liquida è trasferita all'interno di un setaccio vibrante con lo scopo di eliminare impurità e parti granulose. Una volta setacciata, mediante una Pompa a Membrana, viene condotta all'interno di una Filtropressa che separa l'acqua dall'argilla e permette il rassodamento di quest'ultima. Come ultima fase, dopo circa 6/8 ore, l'argilla viene portata alle fasi di impastamento e degassazione nelle quali viene omogeneizzata in assenza di aria.

A questo punto l'argilla è pronta per poter essere trasferita alle fasi successive.

(S.T Levi,2010), (ceramichecear.it).

2.2.2: Foggiatura

La Foggiatura è la prima fase di lavorazione vera e propria della ceramica in cui si determinano le caratteristiche principali della forma finale che si vuole dare alla ceramica lavorata. L'argilla, una volta terminato il processo di preparazione, viene quindi trasferita alle varie lavorazioni che possono essere effettuate manualmente oppure tramite apposite macchine per la foggatura mediante stampaggio.

Tra le principali tecniche di foggatura troviamo:

- Foggiatura manuale;
- Foggiatura per pressatura;
- Foggiatura per colaggio;

La foggatura manuale è un tipo di tecnica di lavorazione dell'argilla effettuate tramite un tornio. Il tornio è un macchinario industriale costituito essenzialmente da un disco che ruota, il cui senso di rotazione è determinato in base alla direzione degli allineamenti obliqui di pori e clasti: se la direzione è a spirale verso l'alto a sinistra, il tornio gira in senso antiorario; viceversa, se la direzione è a spirale verso l'altro a destra, il tornio ruota in senso orario. La foggatura manuale tramite tornio viene utilizzata principalmente per la foggatura di vasellame o di oggetti tondi, proprio perché grazie alla rotazione del disco del tornio e con l'aiuto delle mani di un operatore, si riesce a conferire all'argilla la forma desiderata.

La tecnica di foggatura per pressatura o stampaggio viene utilizzata principalmente per la realizzazione di un elevato numero di pezzi. La forma è ottenuta andando ad inserire l'argilla all'interno di stampi in ceramica, legno o cuoio e successivamente pressata mediante l'utilizzo di presse idrauliche.

Infine, la tecnica di foggatura per colaggio impiega particolari argille caoliniche di colorazione biancastra che vengono prima sciolte in acqua insieme a un fluidificante come il silicato di sodio e, successivamente, vengono appunto colate all'interno di stampi in cui rimangono per circa un'ora. Una volta che l'argilla si è solidificata all'interno dello stampo, quest'ultimo viene tolto e l'oggetto può essere rifinito. (S.T Levi,2010), (ceramichecear.it).

2.2.3: Trattamento delle superfici

La fase di trattamento delle superfici comprende le operazioni di rifinitura, rivestimento e decorazione dell'argilla.

L'operazione di rifinitura si effettua strofinando la superficie con semplici strumenti in modo da ottenere un effetto lucido e liscio del materiale. In questa fase si utilizzano strumenti come ciottoli, stoffe, spatole in legno o in osso, e a volte si applicano anche delle sostanze per ottenere un migliore effetto lucido. In base all'effetto che si vuole ottenere, è possibile distinguere due tipologie di rifinitura: *lisciatura*, per rendere la superficie regolare, liscia e opaca e *lucidatura*, effettuata tramite oggetti rigidi ad essiccamento avanzato per ottenere una parziale impermeabilità del materiale e una maggiore coesione per eventuali rivestimenti o decorazioni. La decorazione dell'argilla si ottiene effettuando delle incisioni sul materiale tramite degli oggetti appuntiti che permettono di creare dei rilievi decorativi. Solitamente le decorazioni sull'argilla sono effettuate manualmente da operatori specializzati, capaci anche di esprimere sullo smalto dell'argilla decori colorati. Per fare ciò si utilizzano dei decori ceramici costituiti da ossidi minerali o minerali misti a fondenti.

La decorazione della ceramica, o meglio la maggiore o minore armonizzazione dei decori realizzati sulla ceramica, e la forma realizzata dell'oggetto, costituiscono due degli elementi più importanti per la valutazione qualitativa dei manufatti. (S.T Levi,2010), (ceramichecear.it).

2.2.4: Fase di cottura

La fase di cottura della ceramica deve essere preceduta da una fase di essiccamento che ha l'obiettivo di eliminare, attraverso l'evaporazione, l'acqua residua all'interno dell'argilla per evitare la rottura dell'oggetto nella fase di cottura. Questo processo può avvenire naturalmente esponendo gli oggetti all'aria o forzando l'essiccazione mediante l'uso di macchinari essiccatori.

La cottura degli oggetti avviene in due fasi: prima cottura e seconda cottura.

- La prima cottura viene effettuata all'interno di Forni alimentati elettricamente o a gas che portano i manufatti ad una temperatura di 1000 gradi centigradi in circa 12 ore. Successivamente vi è il graduale raffreddamento dei materiali che può avvenire naturalmente per dissipazione di calore, o forzatamente in tempi che vanno dalle 24/48 ore in base alle dimensioni dell'oggetto.
- La seconda cottura viene effettuata dopo la fase di decorazione degli oggetti. Questa fase dura dalle 8 alle 10 ore e porta gli oggetti a temperature di circa 920/950 gradi

centigradi. Una volta terminato il processo di raffreddamento dei materiali dalla seconda cottura, gli oggetti sono pronti per poter essere immessi sul mercato. (S.T Levi,2010), (ceramichecear.it).

Capitolo 3: Regole di associazione e Rapidminer

3.1: Il Data Mining

In passato, l'analisi dei dati veniva eseguita per ambienti di grandi dimensioni come Stati e Paesi. Successivamente, con l'emergere delle reti intelligenti, l'analisi dei dati ha trovato un ruolo fondamentale nel riconoscimento dei modelli in ambienti di piccola scala. Uno dei metodi principali per lo studio dell'analisi dei dati per identificare modelli comportamentali è il *data mining*.

Con il termine data mining si intende l'estrazione di informazioni considerate interessanti contenute in una base di dati. (L. Cabibbo, 2014). Lo scopo dell'applicazione delle tecniche di data mining è trovare modelli funzionali nascosti nei dati. Infatti, applicando opportuni algoritmi di data mining è possibile acquisire una nuova conoscenza dei dati.

Il data mining può essere principalmente di due tipi: guidato e non guidato.

Una principale differenza che emerge è che, mentre nel data mining guidato una variabile di classificazione deve essere sempre classificata, stimata o prevista per determinare specificatamente uno schema preciso, nel data mining non guidato non esiste una variabile target che deve essere classificata e il suo scopo è quello di identificare modelli generali che non appartengono ad una particolare variabile. (MA Baherifard et al., Journal Operation and Automation in Power Engineering).

I modelli ottenuti da un'analisi di data mining possono essere di due tipi: descrittivi e predittivi.

- I modelli descrittivi permettono di caratterizzare delle proprietà dei dati;
- I modelli predittivi sono in grado di effettuare delle previsioni future sul comportamento del sistema. (G. Amato, 2019).

Negli ultimi anni sono stati sviluppati numerosi algoritmi di data mining che ricoprono un'ampia classe di problemi. A seconda che il modello ottenuto sia predittivo o descrittivo, è possibile individuare una serie di applicazioni o funzionalità:

- Classificazione (Predittivo);
- Clustering (Descrittivo);
- Regole di associazione (Descrittivo);
- Stima (Predittivo);
- Previsione (predittivo);

(MA Baherifard et al., Journal Operation and Automation in Power Engineering).

La classificazione è il processo di individuazione di un modello che permette di definire delle classi separate (categorie) nei dati presi in esame. Essa si articola in una fase di *training* (addestramento) sui dati che si conoscono per costruire il modello ed una fase di *test* in cui, si effettuano delle previsioni utilizzando il modello prodotto per dedurre la classe di appartenenza di nuovi prodotti sulla base delle loro caratteristiche.

Essenzialmente la classificazione permette di costruire modelli sulla base dei legami ricorrenti sulle variabili esplicative, corrispondenti a osservazioni appartenenti ad una medesima classe; questi legami vengono poi successivamente tradotti in regole di classificazione che permettono di predire la classe di osservazione di cui è noto solo il valore degli attributi esplicativi.

Il Clustering dei dati mira a trovare dei modelli nascosti, cercare delle somiglianze e a prevedere il valore futuro delle serie di dati. Le serie di dati sono lunghe stringhe di numeri che rappresentano diverse caratteristiche di un evento. (MA Baherifard et al., Journal Operation and Automation in Power Engineering).

Il Clustering analizza i dati senza consultare nessuna informazione nota sulla classe ma effettua raggruppamenti sulla base di similarità e omogeneità.

Per stabilire delle omogeneità e similarità occorre stabilire una metrica in grado di esprimere quella che è la distanza tra le unità da analizzare. Quindi verranno predisposti degli input sottoforma di una matrice che riporta tutti gli attributi delle unità sui quali si vuole effettuare la classificazione e il risultato (output) corrisponderà all'assegnazione di ciascuna variabile a un gruppo.

I processi di stima e di previsione permettono, sulla base della costruzione di un modello decisionale, di effettuare previsioni sul comportamento futuro del sistema quando gli viene applicato un nuovo dataset. In particolare, la stima tratta una serie di valori di una variabile continua e sconosciuta, mentre il processo di previsione unisce tutte le tecniche di funzionalità e applicabilità della stima e della classificazione per costruire un modello predittivo più completo e complesso.

Tra tutte le tecniche di Data Mining sopraelencate, questo elaborato tratterà unicamente i processi relativi la costruzione di regole di associazione per la costruzione di modelli decisionali. Infatti, tali regole sono state utilizzate concretamente per effettuare il controllo qualità di alcuni processi e successivamente la pianificazione della produzione futura.

(MA Baherifard et al., Journal Operation and Automation in Power Engineering), (G.Amato,2019), (L.Cabibbo,2014).

3.1.1: Regole di Associazione

Le regole di associazione permettono l'individuazione di correlazioni e associazioni tra diversi insiemi di elementi e spesso vengono ricavate da basi di dati che memorizzano delle transazioni, chiamati appunto database transazionali. I Database transazionali sono degli archivi in cui ogni istanza corrisponde a una transazione che è a sua volta formata da un identificatore e da una serie di *item* (oggetti). (G.Amato, 2019).

Una regola di associazione è un'istruzione di tipo "if-them" in cui una combinazione di condizioni dei dati di input mappa in un determinato aspetto dei valori di output. Esse sono un metodo affidabile per comprendere il comportamento interno del modello e le connessioni tra i dati di input e le previsioni.

Una regola associativa presenta la seguente struttura:

$$\text{CORPO} \longrightarrow \text{TESTA} [\text{SUPPORTO}, \text{CONFIDENZA}]$$

In cui corpo e testa sono item, mentre supporto e confidenza sono due valori percentuali che determinano le caratteristiche della regola.

In particolare, dati due insiemi A e B, tali per cui $A \cap B = \emptyset$ e una transazione T, una regola associativa è un'implicazione del tipo $A \Rightarrow B$ con il seguente significato: se A è contenuto in T, allora B è contenuto in T con una certa probabilità. Questa probabilità è proprio la confidenza della regola associativa. In altre parole, la confidenza misura la significatività della regola, cioè definisce tra tutte le istanze che soddisfano il Corpo (A), quale è la percentuale di quelle che soddisfano anche la Testa (B). (AR Troncoso-Garcia et al., Information Fusion Journal).

Confidenza:

$$\frac{\#\text{transazioni che contengono AUB}}{\#\text{transazioni che contengono A}}$$

È importante sottolineare che la confidenza è un parametro fondamentale per la selezione delle regole associative, in quanto, una volta determinato il suo valore soglia (minimo) si vanno a selezionare solo quelle regole associative che hanno un valore superiore o uguale

a questo valore; queste regole sono chiamate regole associative *forti* e rappresentano quelle più significative per il modello da cui è possibile dedurre e stimare comportamenti futuri. Il supporto, invece, definisce la percentuale delle istanze che soddisfano sia il Corpo (A) che la Testa (B); nello specifico misura quella che è la rilevanza statistica della regola associativa.

Supporto:

$$\frac{\text{\#transazioni che contengono AUB}}{\text{\#transazioni totali}}$$

Come per la confidenza, anche il supporto permette di estrarre le regole associative più interessanti a partire da un valore soglia. Una volta fissato il valore minimo di supporto, si è in grado di selezionare soltanto le regole associative soddisfacenti per il processo di analisi considerato. (G.Amato,2019).

Infatti, molte regole associative non sono forti, quindi non interessanti per il problema, proprio perché non superano le soglie minime di supporto e confidenza prefissate.

Tuttavia, non sempre la generazione di regole associative forti si traduce in regole significative e utili al problema; di conseguenza è utile considerare un particolare indice inteso come coefficiente di correlazione tra due eventi A e B (*lift*) che permette di valutare la significatività di una regola. (G.Amato,2019).

$$l = lift\{A \Rightarrow B\} = conf(A \Rightarrow B) / f(B) = f(A \cup B) / f(A)f(B)$$

- Se $l > 1$ gli eventi A e B sono positivamente correlati;
- Se $l < 1$ gli eventi A e B sono negativamente correlati;
- Se $l = 1$ gli eventi sono indipendenti.

3.1.2: L'Algoritmo Apriori

L'algoritmo apriori è lo strumento che permette di estrarre regole associative forti significative a partire da un insieme di transazioni ed è presente in tutti i programmi di analisi statistica applicate al data mining.

Inizialmente, va a generare un insieme di itemset senza procedere all'esplorazione di tutto lo spazio dei candidati e solo successivamente procede all'estrazione delle regole. L'intero algoritmo si basa su un principio fondamentale secondo cui qualunque sottoinsieme di un itemset frequente, è un itemset frequente.

Tale principio si applica al calcolo degli itemset frequenti:

- Iterativamente si calcolano tutti i k-itemset (itemset di k elementi) da k-1 in poi;
- Si calcola il supporto per ciascuno di questi k-itemset;
- Si eliminano i candidati che hanno un valore di supporto inferiore al valore minimo prefissato;
- Il processo si ferma se non è stato generato nessun k-itemset; in alternativa si pone $k=k+1$ e si itera l'intero processo.

La generazione dei candidati è il primo passo dell'intero algoritmo. Si articola essenzialmente in due fasi: *Join* e *Prune*.

Nel passo Join si parte dall'insieme dei k-itemset frequenti e si genera un insieme dei (k+1)-itemset candidati, nel passo successivo, Prune, si eliminano dall'insieme dei (k+1)-itemset candidati tutti quegli itemset per cui esiste un sottoinsieme k che non appartiene all'insieme degli itemset frequenti. (G.Amato,2019), (myservername.com/apriori-algorithm-data-mining).

Per riassumere, l'algoritmo Apriori si basa sulla seguente regola: se un pattern di lunghezza k non è frequente nel database, anche la sua lunghezza (k+1) non sarà frequente. L'idea che sta alla base di questo algoritmo è quella di generare iterativamente un insieme di pattern candidati di lunghezza (k+1) e di verificare le loro frequenze all'interno del database. Tuttavia, la generazione di un set di candidati è fin troppo costosa, soprattutto quando si ha a che fare con un dataset di grandi dimensioni.

Per superare queste limitazioni e migliorare l'efficienza della scoperta delle regole di associazione, sono stati sviluppati diversi metodi e algoritmi alternativi, come ad esempio l'algoritmo FP-Growth. (Júlio César Borba Nandi, et al., Brasile, 2010).

3.1.3: L'Algoritmo FP-Growth

L'algoritmo FP-Growth rappresenta un algoritmo per l'estrazione di regole associative interessanti ed è utilizzato da numerosi software come, per l'appunto, Rapidminer.

Non è altro che un miglioramento del metodo Apriori in cui viene generato un pattern frequente senza la necessità di generare dei candidati.

L'algoritmo necessita dei due parametri di input di supporto e confidenza. Una volta impostati, il primo passo è assemblare la struttura ad albero FP Tree che memorizzerà i primi set di elementi frequenti. Per fare questo, è necessario leggere il database iniziale due volte: una prima volta vengono individuati gli insiemi degli elementi frequenti (itemset) e i

rispettivi supporti, successivamente gli itemset frequenti con valore di supporto superiori a quello specificato, vengono ordinati in ordine decrescente all'interno di una lista. La seconda lettura del database inizia dove ogni transazione viene letta di nuovo: gli insiemi di elementi poco frequenti di tale transazione sono ignorati, quelli frequenti sono inseriti all'interno di un elenco ordinato in base al valore di supporto degli insiemi di elementi memorizzati nella lista. Al termine della seconda lettura del database, tutte le transazioni verranno processate ed inserite nella struttura ad albero FP. Il primo nodo dell'albero (nodo radice) è sempre NULL. Lo scopo dell'albero FP è quello di estrarre il pattern più frequente. Per ogni itemset presente vengono generate basi di pattern condizionali, che vengono utilizzate per la costruzione di conditional FP tree, che mettono in relazione i percorsi frequenti che si connettono ai nodi corrispondenti all'itemset in questione. Sulla base dell'insieme di N set di elementi frequenti trovati attraverso l'albero FP, l'algoritmo FP-Growth genera le regole di associazione. (Júlio César Borba Nandi, et al., Brasile, 2010).

3.2: Rapidminer

Come già detto, lo scopo principale di questo elaborato e dell'intero progetto di tirocinio è quello di estrarre regole associative significative, a partire da set di dati forniti dall'azienda Matrix, che permettono di effettuare uno studio e un'analisi dell'intero processo di qualità dei prodotti sanitari in ceramica e che successivamente permettono di effettuare delle previsioni sulla pianificazione futura.

Di conseguenza, il passo fondamentale per l'intero studio è senza dubbio l'estrazione di regole associative. Per fare questo è stato utilizzato un particolare software sviluppato da alcuni studiosi nel 2001 a Dortmund in Germania chiamato Rapidminer.

RapidMiner è una piattaforma di Data Science progettata proprio per le aziende al fine di facilitare tutti i processi di analisi predittiva e di modellazione statistica, preelaborazione e visualizzazione dei dati, e quelli di valutazione e distribuzione dei risultati. Questa piattaforma fornisce una guida per progettare ed eseguire flussi di lavoro analitici chiamati Processi. Ciascun processo è caratterizzato da una serie di Operatori in grado di eseguire singole attività all'interno del processo e l'output di ogni operatore costituisce l'input dell'operatore successivo. (Northavati Baharun et al., International Journal of Emerging Technology and Advanced Engineering).

Rapidminer presenta un'interfaccia ben definita e comprensibile anche ai non esperti di programmazione in quanto non richiede la scrittura di alcun codice di programmazione.

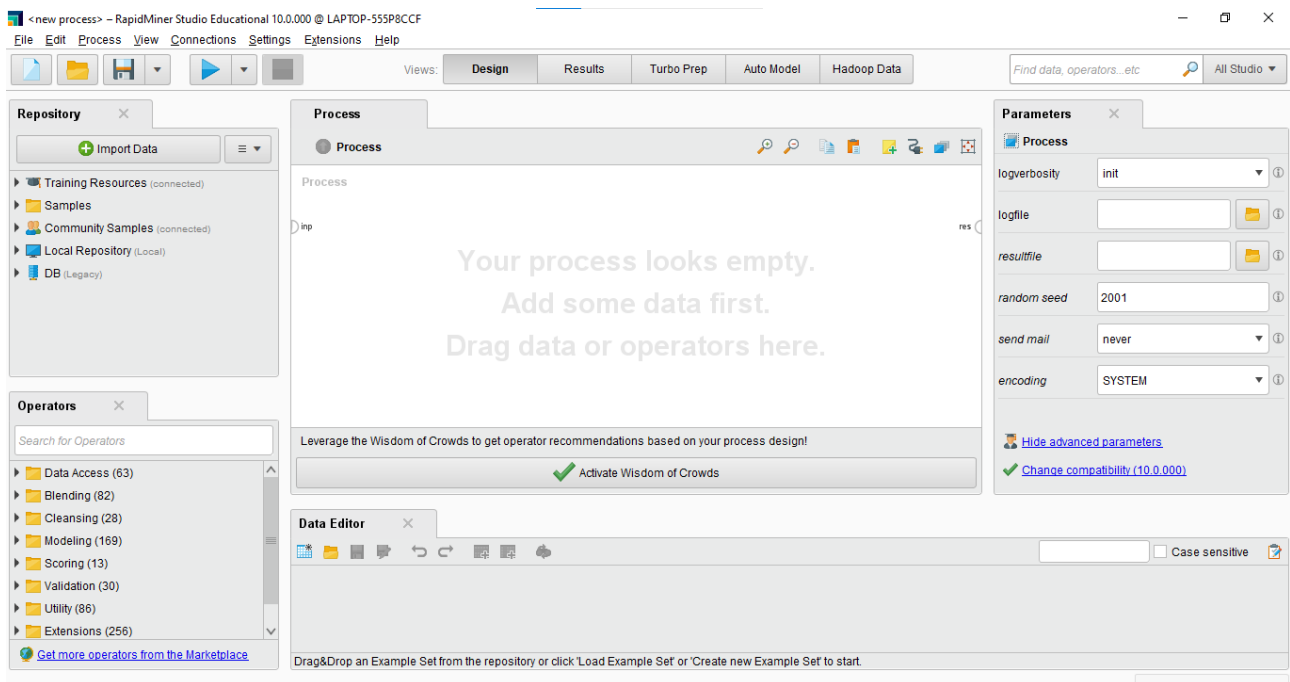


Figura 3: Interfaccia di RapidMiner

Una volta aperto il programma, l'interfaccia di presentazione è quella descritta dalla Figura 3. In alto al centro si trova la finestra di progettazione del processo e quella dei risultati. Entrambe queste finestre possono essere descritte come aree di lavoro che sono attrezzate e ottimizzate per volgere attività specifiche man mano che si intraprende la convalida della modellazione, la preparazione dei dati e l'operatività dei processi di analisi.

Immediatamente sotto si trova la finestra che descrive il processo vero e proprio; come già accennato precedentemente, Rapidminer non prevede la scrittura di un codice, ma il processo è realizzato mediante "blocchetti" collegati gli uni agli altri in modo successivo e collegati alle porte di input/output.

I "blocchetti" che definiscono i processi non sono altro che gli Operatori di Rapidminer. È possibile trovare tutti gli operatori nel pannello in basso a sinistra suddivisi in diverse aree: Data Access, Blending, Cleansing, Modeling, Scoring, Validation, Utility, Extensions e Deployment. Ognuno di questi operatori può essere inizializzato, impostando i parametri che lo caratterizzano, nel pannello dei parametri posto in alto a destra accanto la finestra del processo.

Infine, in alto a sinistra si trova il pannello del Repository attraverso il quale è possibile accedere ai dataset presenti sui documenti del computer e in cui si andranno a salvare tutti i processi realizzati. (rapidminer.com)

3.2.1: Processi realizzati

Con l'aiuto di Rapidminer, durante lo svolgimento del periodo di tirocinio, sono stati realizzati diversi processi per l'estrazione di regole associative. Prima di descrivere nello specifico i processi realizzati in relazione ai dataset utilizzati, è importante descrivere le diverse funzionalità degli operatori inseriti nei processi.

A tal proposito, di seguito, saranno proposte le due tipologie di processi realizzati con annessa spiegazione delle sole funzionalità degli operatori che li costituiscono. Solo successivamente, nel capitolo 4 dell'elaborato, saranno specificati i passaggi della costruzione dei processi in relazione ai dati che l'azienda ha messo a disposizione.

1. La Figura 4 mostra il primo processo realizzato tramite Rapidminer.

Sono stati utilizzati diversi operatori per la sua costruzione, alcuni dei quali saranno riutilizzati per la costruzione del processo successivo.

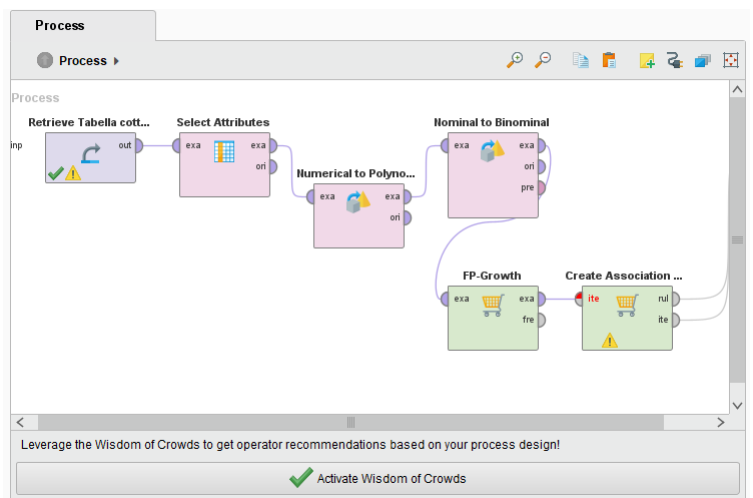


Figura 4: Primo processo realizzato

- Primo blocchetto: Tabella del dataset iniziale su cui si vogliono estrapolare le regole associative;
- Secondo operatore: Select Attributes che permette di selezionare gli attributi su cui si vogliono costruire le regole associative. Fornisce inoltre anche diversi tipi di filtro per semplificare la selezione degli attributi, come la selezione diretta, tramite un'espressione regolare o selezionando solo gli attributi senza valori mancanti. Una particolarità di questo operatore è che solo gli attributi selezionati vengono trasferiti alla porta di output e consegnati all'operatore successivo.
- Terzo operatore: Numerical to Polynominal. Questo operatore modifica il tipo di attributi numerici selezionati in un tipo polinomiale. Non solo modifica il tipo di attributi selezionati, ma mappa anche tutti i valori di questi attributi ai corrispondenti valori polinomiali. In altre parole, ogni valore numerico viene

semplicemente utilizzato come valore nominale del nuovo attributo. Poiché gli attributi numerici possono avere un numero enorme di valori diversi anche in intervalli ridotti, la conversione di un attributo numerico in forma polinomiale genererà un numero enorme di possibili valori per il nuovo attributo.

- Quarto operatore: Nominal to Binominal. Permette di modificare non solo il tipo di attributi selezionati (da nominale a binominali), ma anche di mapparne il valore. Essendo attributi binominali, quindi binari, possono assumere soltanto due valori: vero o falso.
- Quinto operatore: FP-Growth. L'FP-Growth è un vero e proprio algoritmo che permette di selezionare gli elementi che ricorrono frequentemente in un database transazionale. È l'operatore che permette di cominciare ad estrarre regole associative forti. Infatti, tra i parametri da inserire per inizializzare questo operatore vi è proprio il parametro di supporto minimo che, come si è definito precedentemente, è uno dei due parametri essenziali per l'estrazione di regole significative.
- Ultimo operatore: Create Association Rules. Permette la generazione di regole associative dal dataset di elementi frequenti specificato dall'FP-Growth. Quando si inserisce questo operatore è importante settare il valore di confidenza minima per ottenere regole associative interessanti.

2. La Figura 5 mostra il secondo processo realizzato tramite Rapidminer.

Come si può notare, molti degli operatori che lo costituiscono sono gli stessi utilizzati nel processo precedente e le loro funzionalità sono praticamente le stesse.

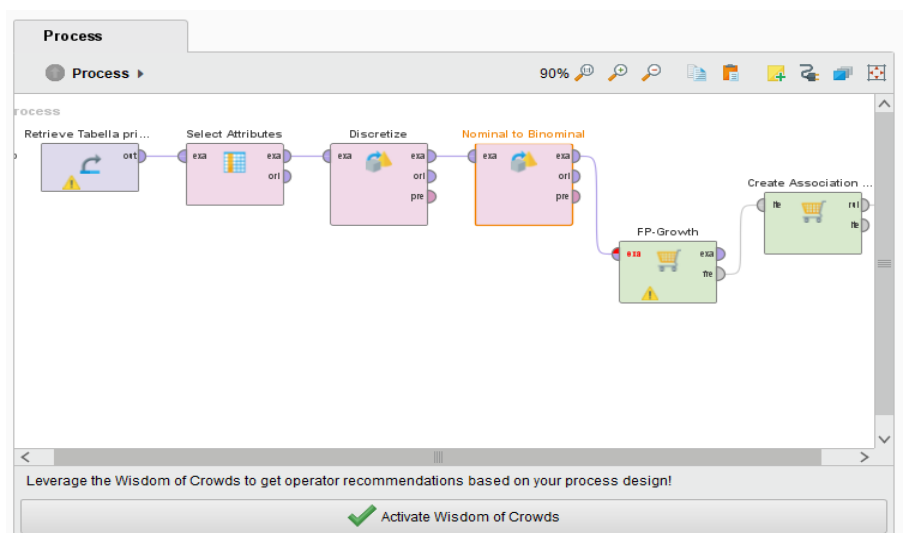


Figura 5: Secondo processo realizzato

L'unico operatore che differenzia i due processi è: Discretize by Biginning.

- Discretize by Biginning: questo operatore discretizza gli attributi numerici selezionati in attributi nominali. Il parametro *number of bins*

viene specificato per determinare il numero di bin, cioè il numero di intervalli che si vuole creare. Gli intervalli di valori numerici sono suddivisi in segmenti di uguale dimensione e ogni segmento rappresenta un contenitore a cui sono assegnati dei valori. Ogni intervallo viene definito automaticamente e il formato di denominazione può essere modificato utilizzando il parametro del tipo di nome dell'intervallo. Inoltre, i parametri *min value* e *max value* vengono utilizzati per definire i limiti dell'intervallo. Infatti, se sono presenti valori inferiori o superiori al parametro di valore, viene creato un intervallo separato.

(rapidminer.com).

CAPITOLO 4: Dataset e analisi dei risultati

In questo capitolo, viene presentato il progetto svolto in collaborazione con l'azienda Matrix Modellerie. L'azienda ha fornito una serie di documenti, per un periodo che va da Giugno 2021 a Giugno 2022, in forma grezza redatti dai dipendenti impiegati nel reparto del controllo qualità. Questi documenti sono stati dapprima convertiti in formato excel e successivamente utilizzati come dataset di partenza per generare l'intero processo di analisi e di estrazione delle regole associative sul software Rapidminer.

4.1: Dataset iniziale

L'azienda Matrix sottopone, giorno per giorno, due diverse tipologie di impasti della ceramica al controllo qualità. Come si è dimostrato precedentemente, è fondamentale effettuare controlli giornalieri sulla qualità delle materie prime utilizzate per evitare, in fase di produzione, di realizzare prodotti non conformi alle specifiche definite. Effettivamente, avviare un ciclo produttivo con materie prime non idonee comporterà, oltre ad uno spreco di tempo, anche elevati costi per un eventuale smaltimento del prodotto o per una sua rilavorazione.

Di conseguenza, gli operatori impiegati al controllo qualità degli impasti, hanno il compito giornaliero di redigere su un modulo tutte le caratteristiche considerate fondamentali per ottenere un impasto idoneo e utilizzabile in fase di produzione.

Gli impasti su cui l'azienda effettua il controllo qualità sono l'FFC Imerys e il VC Euroarce e le caratteristiche di cui gli operatori sono tenuti a tenere traccia sono:

- Peso litro: solitamente si aggira intorno i 1835 per l'impasto VC Euroarce e intorno i 1945 per l'impasto FFC Imerys;
- Viscosità: intorno i 270 per l'impasto FFC Imerys e tra i 295 e i 300 per l'impasto VC Euroarce;
- Tixotropia 1 min: per entrambi gli impasti è solitamente compresa tra 20/25;
- Tixotropia 6 min: per il VC Euroarce è compresa tra i 65/75, mentre per l'altro impasto ha solitamente valori che vanno da 90 a 100;
- Temperatura in °C.

Oltre queste caratteristiche, gli operatori impiegati nel controllo qualità devono specificare il giorno in cui sono stati effettuati i controlli, l'orario, eventuali note aggiuntive e ovviamente il loro nome.

Da questi moduli riempiti in azienda all'interno del reparto, parte tutta l'attività di tirocinio effettuata.

Inizialmente questi moduli sono stati convertiti in file Excel e successivamente integrati con altri documenti contenenti il controllo qualità dei prodotti finiti. Nello specifico, in corrispondenza dei dati sul controllo qualità dei prodotti finiti, sono stati inseriti i dati relativi al controllo qualità degli impasti con cui questi ultimi sono stati realizzati.

Successivamente, il file Excel completo, contenente cioè i dati sul controllo qualità dei prodotti e degli impasti sulle prime cotture e sulle cotture di ripasso, è stato utilizzato come dataset iniziale per effettuare tutto lo studio sull'estrazione delle regole associative su Rapidminer.

4.2: Estrazione delle regole e considerazioni

La struttura dei processi realizzati su Rapidminer è stata ampiamente illustrata nel capitolo 3 dell'elaborato. A questo punto, saranno illustrati dettagliatamente i processi applicati ai dati a disposizione.

A ciascuno dei due processi realizzati (Figura 4 e Figura 5) sono stati applicati due differenti tipi di dataset iniziali: quello relativo alle prime cotture e quello sulle cotture di ripasso. Di conseguenza si avranno due processi differenti (per ciascuno di quelli realizzati) identici nella struttura ma diversi nei risultati e nelle regole estratte.

4.2.1: Regole associative primo processo

Il primo processo, visibile nella Figura 4, estrae le regole associative mettendo in relazione soltanto le variabili inerenti all'articolo, la quantità da scartare, la quantità da ripassare e la quantità di prodotti buoni realizzati.

Per estrarre regole che considerano soltanto queste variabili, basta semplicemente inserire nel processo di Rapidminer l'operatore Select Attributes e settarlo andando a trascinare a destra gli attributi d'interesse (Figura 6).

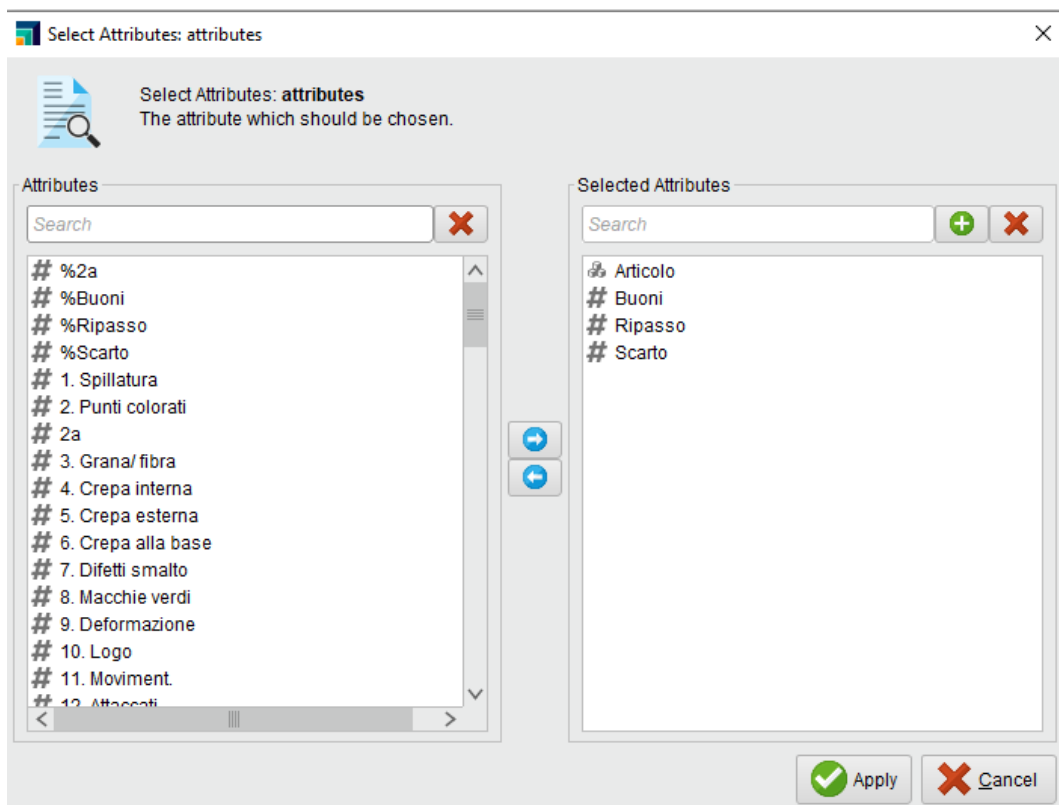


Figura 6: Operatore Select Attributes in Rapidminer

Gli attributi settati nell'operatore Select Attributes sono trasferiti agli operatori successivi che permettono l'estrazione delle regole. In particolare, l'operatore FP-Growth setta il valore di supporto minimo, impostato a 0.01, mentre l'operatore Create Association Rules setta il valore di confidenza impostato a 0.05.

Una volta impostati e settati tutti gli operatori del processo, è possibile mandarlo in esecuzione tramite il tasto posto in alto a sinistra dell'interfaccia.

Il processo fornirà una serie di regole associative del tipo:

NOME ARTICOLO= BUONI, RIPASSO o SCARTO [supporto, confidenza]

Tra le regole generate dal processo sono da considerare interessanti solo quelle per cui i valori di supporto e confidenza sono molto maggiori rispetto quelli minimi impostati.

La Figura 7 mostra alcune delle regole generate dal primo processo. Ad esempio, la regola descritta dalla riga 179: **Articolo= Clo 715 ⇒ Buoni=1 [0.0027, 0.21]** può essere tradotta nel seguente modo: dell'articolo Clo 715 è stato prodotto un quantitativo pari ad 1 di pezzi buoni, cioè di pezzi che non hanno bisogno di essere ripassati o scartati e che quindi possono essere immessi direttamente sul mercato. Questa regola può essere considerata

significativa in quanto presenta un valore di confidenza pari a 0.21 che è molto maggiore rispetto il valore minimo fissato pari a 0.05.

Premises	Conclusion	Support	Confidence
Articolo = Clo 715	Buoni = 1	0.0027033289565150226	0.21084337349397592
Articolo = Juno Wc	Buoni = 2	0.001081331582606009	0.2153846153846154
Articolo = Biis	Ripasso = 1	0.004016374449679463	0.21666666666666667
Articolo = ETRE 33	Buoni = 1	0.0013902834633505831	0.21686746987951808
Articolo = ETRE 33	Buoni = 2	0.0013902834633505831	0.21686746987951808
Articolo = Apron 30"	Scarto = 1	0.0013130454931644397	0.23611111111111111

Figura 7: Esempi di regole associative del primo processo

Tutte le regole associative generate dal primo processo sono state poi riportate su un file Excel in cui sono state evidenziate quelle di maggior interesse. Questo primo processo, tuttavia, genera delle regole da cui non è possibile effettuare previsioni future sulla produzione ma semplicemente descrive il quantitativo prodotto di ogni articolo e ne evidenzia lo stato all'interno del ciclo produttivo in quanto non sono state considerate particolari caratteristiche di analisi (cosa che invece verrà fatta nel processo successivo).

4.2.2: Regole associative secondo processo

Il secondo processo, descritto nella Figura 5, ha l'obiettivo di estrarre regole associative che permettono di effettuare specifiche analisi di previsione sulla pianificazione di futuri cicli di produzione. Sono stati realizzati due processi identici con l'unica differenza che uno considera tutte le caratteristiche dei pezzi a $t = -4$, l'altro le considera a $t = -5$.

Tali processi risultano essere più complessi rispetto al precedente in quanto oltre a prendere in considerazione le variabili di articolo, %buoni, %ripasso, %scarto, considerano anche diverse caratteristiche specifiche degli oggetti realizzati.

Nello specifico, per effettuare la corretta estrazione delle regole, è necessario settare l'operatore Select Attributes che permette di mettere in relazione l'articolo con le seguenti caratteristiche scelte: spillatura, punti colorati, grana/fibra, crepa interna, crepa esterna, crepa alla base, difetti smalto, macchie verdi, deformazione, logo, attaccati, bolle, °C, peso

litro, pressione, punto di rugiada, tixotropia 1min e 6min, Tmax °C, Tmin °C, Tmedia °C, umidità, vento max km/h, vento medio km/h, viscosità e visibilità.

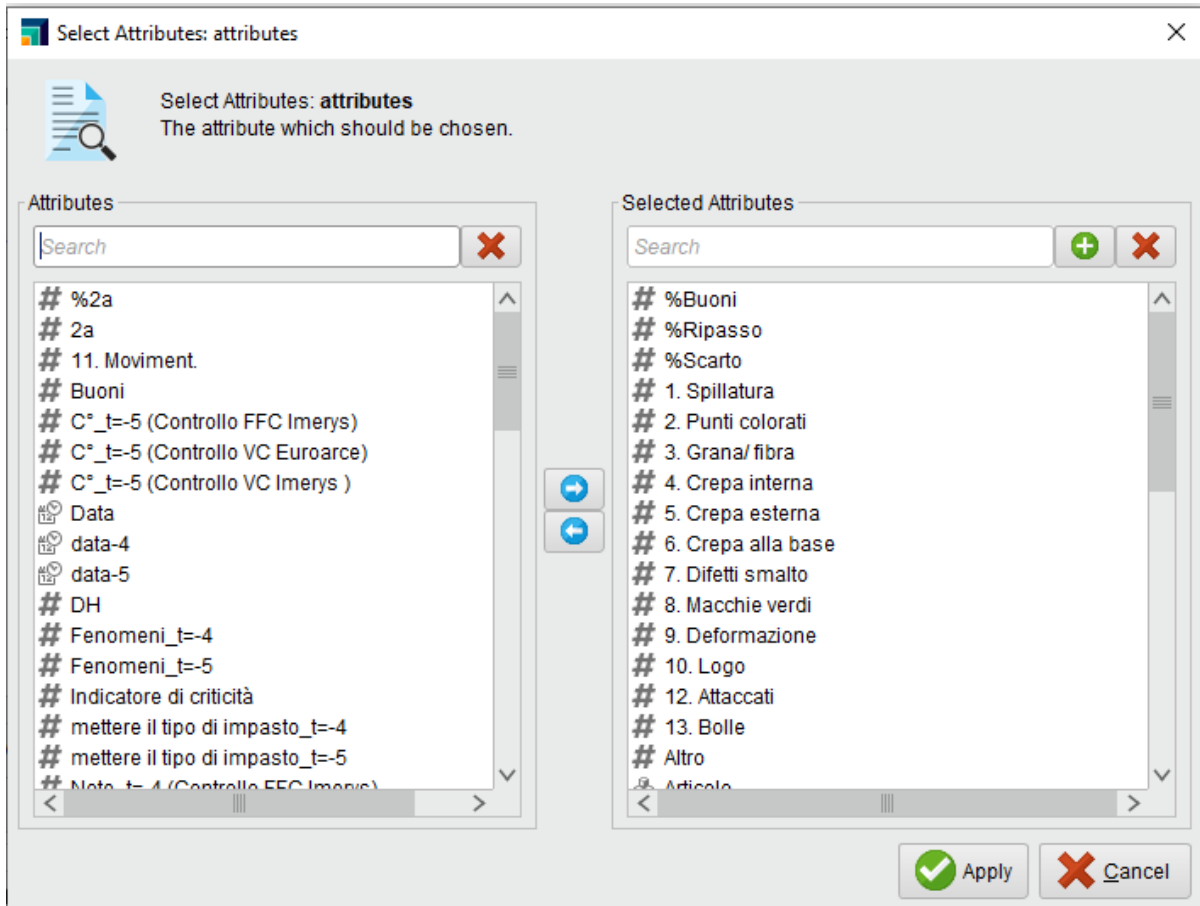


Figura 8: Select Attributes secondo processo

Esattamente come il processo precedente, per settare l'operatore Select Attributes si trascina a destra tutti gli attributi che si vogliono mettere in relazione per estrarre le regole associative.

Particolarmente rilevante nel processo è l'utilizzo dell'operatore Discretize. Come precedentemente illustrato, tale operatore permette di discretizzare gli attributi andando a creare dei range (intervalli) in cui un certo comportamento del processo si ripete con continuità. Mandando in esecuzione il processo, vengono estratte regole associative della seguente forma:

56	Articolo = Orchard 36 "	%Scarto = range1 [-∞ - 0.020]	0.00432532	0.71794
0			633042403	8717948
			6	718

56 1	Articolo = Orchard 36 "	%Buoni = range10 [0.900 - 1]	0.00432532 633042403 6	0.71794 8717948 718
56 2	%Scarto = range1 [-∞ - 0.020], Articolo = Apron 33"	%Buoni = range10 [0.900 - 1]	0.00610179 964470533 75	0.71818 1818181 8181
56 3	%Ripasso = range1 [-∞ - 0.040], Articolo = Apron 36" SB	%Scarto = range1 [-∞ - 0.020]	0.00278056 692670116 63	0.72
56 4	Articolo = Apron 30"	%Buoni = range10 [0.900 - 1]	0.00401637 444967946 3	0.72222 2222222 2222
56 5	%Buoni = range1 [-∞ - 0.100], Articolo = Juno Wc	%Ripasso = range25 [0.960 - 1]	0.00200818 722483973 13	0.72222 2222222 2222
56 6	Articolo = Juno bidet	%Scarto = range1 [-∞ - 0.020]	0.00262609 098632887 94	0.72340 4255319 149

Figura 9: Esempi regole associative secondo processo

Tuttavia, non tutte le regole estratte sono significative per lo scopo dell'analisi. Maggiormente significative sono quelle che presentano le seguenti strutture:

- Articolo \Rightarrow %scarto = range n.
- Articolo \Rightarrow %ripasso = range n.

Di entrambe le strutture vanno considerate soltanto quelle regole per cui i valori di confidenza sono molto maggiori rispetto i valori minimi.

Anche per questo secondo processo, i valori di supporto e confidenza minimi sono stati impostati rispettivamente a 0.01 e 0.05 in quanto rappresentano i valori più adatti per estrarre un numero di regole significative; infatti, un numero esiguo di regole può risultare insufficiente al modello per apprendere in maniera corretta il fenomeno in analisi. Viceversa, avere un numero troppo elevato di regole può portare il modello ad avere dimensioni troppo elevate e di conseguenza non si è in grado di valutare eventuali comportamenti ricorrenti.

Dalle due strutture di regole significative, è possibile effettuare alcune considerazioni: innanzitutto avere un'elevata percentuale di scarto di alcuni prodotti sotto determinate condizioni (ad esempio ad una certa temperatura) potrebbe portare ad avere mancanza di pezzi a fine ciclo produttivo. Di conseguenza una corretta ipotesi sarebbe quella di aumentare la produzione di tutti quegli articoli che presentano una % di scarto non indifferente, proprio per evitare di non riuscire ad evadere tutti gli ordini dei clienti.

Successivamente, la seconda tipologia di regola associativa mette in evidenza i range per cui un determinato articolo necessita di essere rilavorato a causa di eventuali caratteristiche che non soddisfano quelle necessarie alla sua immissione sul mercato (difetti). Da questa regola è possibile dedurre la quantità da sovra produrre, effettuando la somma delle quantità attese pesate sulle confidenze, per evitare di avere ritardi nella consegna degli ordini.

CONCLUSIONI

Attraverso questo progetto di tesi si è cercato di descrivere e analizzare le principali tecniche statistiche che hanno come obiettivo l'estrazione di regole associative nascoste all'interno di un dataset contenente informazioni grezze. Nello specifico queste tecniche sono state applicate alle informazioni provenienti dal reparto controllo qualità dell'azienda Matrix, azienda con cui è stato svolto un periodo di tirocinio nel periodo da ottobre 2022 a gennaio 2023.

L'intero elaborato si focalizza su un concetto fondamentale: la qualità intesa come progetto di sviluppo e miglioramento continuo per le aziende.

Negli ultimi anni, grazie all'introduzione di numerosi software, come in questo caso Rapidminer, gli studi sulla qualità dei prodotti di un'azienda sono ormai diventati elemento cardine per tutta la parte di pianificazione: a partire dalla pianificazione della produzione fino ad arrivare alla pianificazione di ordini di materie prime da effettuare o addirittura alla pianificazione dell'intero ambiente lavorativo. (D. Grisot,2006).

Nello specifico, si è entrati nel dettaglio della costruzione di modelli/processi all'interno di Rapidminer. Sono stati costruiti diversi processi e da questi sono state effettuate analisi che hanno permesso di pianificare la produzione futura.

Tutte le aziende che utilizzano queste nuove tecniche di data mining e analisi dei processi, hanno raggiunto o continuano a raggiungere l'obiettivo di miglioramento continuo, a partire dall'ottimizzazione dei tempi con cui viene effettuato il controllo qualità. Come già ampiamente illustrato, inizialmente il controllo qualità era una pratica svolta unicamente visivamente dagli operatori impiegati nel reparto e di conseguenza i tempi erano nettamente più lunghi e i risultati meno attendibili.

Tali tecniche sono impiegate da molte imprese di qualsiasi settore, a partire dal settore manifatturiero fino ad arrivare al settore della moda. Sono infatti tecniche universali applicabili in qualsiasi ambito, da grandi e piccole imprese, e se applicate nel modo corretto portano a risultati univoci e attendibili.

L'attuale Era dell'*Industria 4.0* ha diffuso la consapevolezza, tra le aziende, che il focus centrale debba essere il cliente e la sua soddisfazione e di conseguenza bisogna eliminare tutte quelle attività che provocano sprechi e che non aggiungono valore al cliente. Proprio per questo motivo, le imprese effettuano investimenti sulle moderne tecnologie di analisi come supporto al lavoro umano. (M.Bevilacqua, 2017/2018).

BIBLIOGRAFIA

- D. Grisot, “La gestione della qualità- capire e applicare la norma ISO 90001”, 2006 Tecniche Nuove, Milano.
- F. Marmura, “Qualità e servizio bancario: un’analisi”, Gennaio 2012.
- A.V. Feigenbaum, “Total Quality Control”, New York, MC Graw-Hill, 1961.
- J. Juran, “La qualità nella storia”, 1997.
- P. Senni, “La filosofia di Deming e il ciclo PDCA”.
- E. Deming, “The essential Deming: Leadership Principles from the father of Quality”, edizione 2012.
- G. Bizzarri, C. Guarnieri, “Gli otto principi per la qualità: l’approccio e la gestione dei processi”.
- Le Yao, Zhiquiange, “Casual variable selection for industrial process quality prediction via attention- based GRUNetwork”- Engineering Application of Artificial Intelligence
- M. Chaabi, M. Hamlich, M. Garounai, “Product defected detection based on convolucional autoencored and one-class classification”- IAES International Journal of Artificial Intelligence (IJ-AI)
- M. A. Baherifard, R. Kazemzadeh, A.S. Yazdankhah , M. Marzband- “Improving the Effect of Electric Vehicle Charging on Imbalance Index in the Unbalanced Distribution Network Using Demand Response Considering Data Mining Techniques”- Journal of Operation and Automation in Power Engineering.
- Norhayati Baharun , Nor Faezah Mohamad Razi , Suraya Masrom , Nor Ain Mohamad Yusri , Abdullah Sani Abd Rahman -“Auto Modelling for Machine Learning: A Comparison Implementation between RapidMiner and Python”- International Journal of Emerging Technology and Advanced Engineering.
- A.R Troncoso-Garcia, M. Martinez-Ballesteros, F. Martinez-Alvarez “A new approach based on association rules to add explainability to time series forecasting models”- Information Fusion Journal.
- S.T Levi, “Dal coccio al vasaio. Manifattura, tecnologia e classificazione della ceramica.” -2010.
- Luca Cabibbo, Riccardo Torlone, “Introduzione al Data Mining” -settembre 2014.
- G. Amato, “Analisi dei dati ed estrazione della conoscenza”- Corso di Laurea Specialistica in Economia Informatica, Università “G. D’Annunzio”, Chieti-Pescara, marzo 2019.

- G. Amato, “Regole Associative” - Corso di Laurea Specialistica in Economia Informatica, Università “G. D’Annunzio”, Chieti-Pescara, marzo 2019.
- Júlio César Borba Nandi, et al.- “O Algoritmo de Associação Frequent Pattern-Growth na Shell Orion Data Mining Engine”- Universidade do Extremo Sul Catarinense (UNESC)- Brasile, 2010.
- L. Screpanti, “Machine Learning” -Corso di Modellistica e ottimizzazione per i processi industriali, Ingegneria Gestionale UNIVPM, marzo 2022.
- L. Screpanti, “Approccio Predittivo” - Corso di Modellistica e ottimizzazione per i processi industriali, Ingegneria Gestionale UNIVPM, marzo 2022.
- M. Bevilacqua, “Introduzione a INDUSTRY 4.0” -Corso di Logistica Industriale, Ingegneria Gestionale UNIVPM, 2017/2018.

SITOGRAFIA

- <https://www.qualitiamo.com/guru/juran/introduzione.html>
- <https://qualitiamo.com/guru/juran/trilogia.html>
- <https://qualitiamo.com/guru/deming/introduzione.html>
- <https://ita.myservername.com/apriori-algorithm-data-mining>
- <https://ceramichecear.it/pages/lavorazione-della-ceramica>
- <https://www.matrixmodelleria.com/azienda.html>
- <https://www.etoceramiche.it/azienda/>
- <https://www-scopus-com.ezproxy.cad.univpm.it/results/results.uri?sort=plf-f&src=s&st1=rapidminer&nlo=&nlr=&nls=&sid=006007cd81aa9c3e0a>
- https://scholar.google.it/schhp?hl=it&as_sdt=0,5
- <https://rapidminer.com/>

RINGRAZIAMENTI