

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

RIEPILOGO SCHEDA DI CASO

RISULTATO ATTESO 1 - Eseguire operazioni di raccolta dati e data mining, provvedendo alla preparazione per i successivi trattamenti

CASI ESEMPLIFICATIVI:

Dimensione 1 - Raccolta e Preparazione Dati: **4 casi**

Dimensione 2 - Operazioni di Data mining: **3 casi**

Dimensione 3 - Consolidamento Dati: **1 caso**

RISORSE A SUPPORTO DELLA VALUTAZIONE (RSV)

RISULTATO ATTESO 2 - Realizzazione di analisi avanzate di data set (anche di notevoli dimensioni e di big data), anche a partire dalla creazione di modelli, algoritmi e strumenti di raccolta e analisi dati, provvedendo, successivamente, a realizzare rapporti per supportare i processi decisionali

CASI ESEMPLIFICATIVI:

Dimensione 1 - Procedure data curation and data quality control: **4 casi**

Dimensione 2 - Creazione di modelli, algoritmi e strumenti per la raccolta e analisi dei dati: **4 casi**

Dimensione 3 - Cura dei rapporti di sintesi e di comunicazione dell'esito dell'analisi a supporto delle decisioni: **5 casi**

RISORSE A SUPPORTO DELLA VALUTAZIONE (RSV)

RISULTATO ATTESO 3 - Pianificare e coordinare i processi di data analytics, assicurando il corretto feeding dei livelli strategici e decisionali

CASI ESEMPLIFICATIVI:

Dimensione 1 - Pianificazione e coordinamento dei processi: **3 casi**

Dimensione 2 - Trasferimento alle strategie di business: **4 casi**

RISORSE A SUPPORTO DELLA VALUTAZIONE (RSV)

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

SCHEDA DI CASO

RISULTATO ATTESO 1 - Eseguire operazioni di raccolta dati e data mining, provvedendo alla preparazione per i successivi trattamenti

1 - RACCOLTA E PREPARAZIONE DATI

Grado di complessità 4

1.4 DATA PREPARATION

Promuovere operazioni di data cleaning, eliminando possibili errori e stabilendo dei meccanismi di comportamento in caso di dati mancanti, e di pre-processing, esplorando il target data ed effettuando un'analisi su base campionaria

Grado di complessità 3

1.3 SELEZIONE DI UN TARGET DATA SET

Selezionare, decodificare e segmentare i dati grezzi secondo criteri prestabiliti al fine di pervenire ad un sottoinsieme di variabili e di dati o di un campione di dati (structured data, unstructured data) che rappresentano un determinato target data o dati obiettivo

Grado di complessità 2

1.2 TRANSFORMATION & CONFIGURATION

Effettuare operazioni di trasformation, convertendo tipi di dati in altri o definendo nuovi dati ottenuti attraverso l'uso di operazioni matematiche o logiche sulle variabili, e configuration, effettuando una loro riconfigurazione in caso di provenienza da fonti diverse

Grado di complessità 1

1.1 DATA GATHERING

Raccogliere dati e informazioni "grezze", disponibili in diversi format (.pdf, .csv, .txt, ecc.), utilizzando fonti attendibili, certificate e coerenti con il fabbisogno richiesto

2 - OPERAZIONI DI DATA MINING

Grado di complessità 2

2.2 SCELTA DEL PROCESSO DI DATA MINING

Stabilire la tipologia di tecnica di data mining da utilizzare (cluster analysis, regression analysis, classification analysis, anomaly detection analysis, intrusion detection, association rule learning,

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

decision tree, neural networks, rule induction, data warehouse) più idonea rispetto all'obiettivo dell'analisi

2.2 SCELTA DEGLI ALGORITMI DI DATA MINING

Selezionare il metodo da usare per ricercare patterns nei dati, definendo quali parametri possono essere più appropriati e integrando i metodi di data mining scelti con i criteri generali di Data Analytics (Knowledge Discovery in Databases)

Grado di complessità 1

2.1 DATA MINING

Adattare l'algoritmo standard prodotto al caso preso in esame, in base alla tipologia di data mining selezionata (cluster analysis, regression analysis, classification analysis, ecc.), applicandolo ad relativo database e fornendo come outcome la descrizione dell'informazione ricercata

3 - CONSOLIDAMENTO DATI

Grado di complessità 1

3.1 CONSOLIDAMENTO DEI MODELLI IDENTIFICATI

Analizzare e verificare i risultati ottenuti e le performance del sistema, anche valutando possibile retroazione a fasi precedenti, al fine di migliorare l'efficacia dei modelli di data analytics

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

SCHEDA RISORSE A SUPPORTO DELLA VALUTAZIONE DEL RISULTATO ATTESO 1

RISORSE FISICHE ED INFORMATIVE TIPICHE (IN INPUT E/O PROCESS ALLE ATTIVITÀ)

- Obiettivi di analisi
- Norme in materia di protezione della privacy e protezione dati personali (GDPR)
- Criteri generali di Data Analytics (Knowledge Discovery in Databases)
- Data Analytics (Alteryx, Apache Spark, SAS, Big ML, MATLAB, Jupyter)
- Structured data
- Unstructured data
- Dati grezzi (.pdf, .csv, .txt)

TECNICHE TIPICHE DI REALIZZAZIONE/CONDUZIONE DELLE ATTIVITÀ

- Tecniche di data cleaning
- Tecniche di data processing
- Tecniche di data transformation
- Tecniche di data configuration
- Tecniche di data gathering
- Cluster analysis
- Regression analysis
- Classification analysis
- Anomaly detection analysis
- Intrusion detection analysis
- Association rule learning
- Decision tree
- Neural networks
- Rule induction
- Data warehouse

OUTPUT TIPICI DELLE ATTIVITÀ

- Raccolta e preparazione dati effettuata (data cleaning, data transformation e configuration, data catering)
- Algoritmo di Data Mining selezionato e adattato
- Descrizione dell'informazione ricercata

INDICAZIONI A SUPPORTO DELLA SCELTA DEL METODO VALUTATIVO E DELLA PREDISPOSIZIONE DELLE PROVE

ESTENSIONE SUGGERITA DI VARIETÀ PRESTAZIONALE

1. L'insieme delle tecniche e metodologie di data gathering e preparation
2. L'insieme delle tecniche e metodologie di data transformation e configuration

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

3. L'insieme di regole relative alla data mining
4. Un set di obiettivi di analisi
5. Un set di dati grezzi

DISEGNO TIPO DELLA VALUTAZIONE

1. Prova prestazionale: a partire dai set dati, simulazione delle operazioni di raccolta e preparazione dati (data gathering, data preparation, data cleaning, data transformation & configuration) e data mining (con scelta di una delle tecniche a disposizione)
2. Colloquio tecnico relativo alle scelte adottate nella raccolta e preparazione dei dati e nell'applicazione dell'algoritmo standard per le operazioni di data mining

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

SCHEDA DI CASO

RISULTATO ATTESO 2 - Realizzazione di analisi avanzate di data set (anche di notevoli dimensioni e di big data), anche a partire dalla creazione di modelli, algoritmi e strumenti di raccolta e analisi dati, provvedendo, successivamente, a realizzare rapporti per supportare i processi decisionali

1 - PROCEDURE DATA CURATION AND DATA QUALITY CONTROL

Grado di complessità 4

1.4 EXPLORATION

Assicurare una maggiore comprensione analitica dei dati raccolti in funzione degli obiettivi delle analisi, in modo da avere una caratterizzazione più accurata delle informazioni disponibili

Grado di complessità 3

1.3 CLEANING

Individuare e correggere i dati grezzi (raw data) provenienti da fonti diversi che contengono imprecisioni, armonizzando le relative informazioni mediante specifici dataset (es. Alteryx, Apache Spark, SAS, Big ML, MATLAB, Jupyter)

Grado di complessità 2

1.2 ARRICCHIMENTO DEL DATABASE

Categorizzare i dati disponibili a seconda del loro uso finale creando nuove variabili dai valori originali, al fine di consentire un loro facile utilizzo

Grado di complessità 1

1.1 STANDARDIZZAZIONE

Promuovere una standardizzazione del processo di data curation, attraverso l'utilizzo di processi semi-automatici (ontologies) di text mining per identificare eventuali sovrapposizioni di metadati

2 - CREAZIONE DI MODELLI, ALGORITMI E STRUMENTI PER LA RACCOLTA E ANALISI DEI DATI

Grado di complessità 4

2.4 CLASSIFICATION MODELING

Assegnare ogni campione alla giusta categoria, avendo a disposizione una serie di dati appartenenti

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

ad un numero finito di categorie note

Grado di complessità 3

2.3 CLUSTERING MODELING

Creare i gruppi in modo tale che gli elementi al loro interno abbiano dei punti in comune, avendo a disposizione una serie di dati appartenenti ad un numero finito di gruppi non noti

Grado di complessità 2

2.2 NATURAL LANGUAGE PROCESSING

Sfruttare le potenzialità dei modelli di elaborazione del linguaggio naturale (NLP) per il riconoscimento e l'estrazione automatica di informazioni

Grado di complessità 1

2.1 FORECASTING MODELING

Esaminare le caratteristiche chiave di una serie ordinata di campioni, per rilevare la stagionalità delle tendenze, l'influenza di fattori esterni, e prevedere i valori futuri di una variabile in base alla sua relazione con le altre o con determinate serie storiche

3 - CURA DEI RAPPORTI DI SINTESI E DI COMUNICAZIONE DELL'ESITO DELL'ANALISI A SUPPORTO DELLE DECISIONI

Grado di complessità 5

3.5 DATA VISUALISATION

Estrarre e interpretare le informazioni mediante l'utilizzo di strumenti di data visualisation (es. Tableau, RawGraphs, ChartBlocks, PowerBI, QlikSense, ecc.), così da comunicare i risultati con i propri stakeholders in modo efficace

Grado di complessità 4

3.4 DESCRIPTIVE ANALYTICS

Predisporre analisi "descrittive" con riferimento a dati riguardanti situazioni attuali o precedenti relativi al mercato, identificando i principali KPI e gli indicatori di prestazione

Grado di complessità 3

3.3 PREDICTIVE ANALYTICS

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

Predisporre analisi “predittive”, mediante l'utilizzo di tool programmati per fornire risposte relative alla tendenze del mercato nel futuro e avvalendosi di strumenti matematici (forecasting; regressione, applicazione modelli predittivi)

Grado di complessità 2

3.2 PRESCRIPTIVE ANALYTICS

Predisporre analisi “prescrittive”, proponendo soluzioni operative/strategiche, sulla base delle analisi svolte, fra le quali il decision maker potrà scegliere

Grado di complessità 1

3.1 AUTOMATED ANALYTICS

Predisporre analisi “automatizzate”, capaci di mettere in atto autonomamente le scelte che, in base alle analisi svolte, ritengono valide rispetto agli obiettivi prefissati

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

SCHEDA RISORSE A SUPPORTO DELLA VALUTAZIONE DEL RISULTATO ATTESO 2

RISORSE FISICHE ED INFORMATIVE TIPICHE (IN INPUT E/O PROCESS ALLE ATTIVITÀ)

- Obiettivi di analisi
- Norme in materia di protezione della privacy e protezione dati personali (GDPR)
- Linguaggio di programmazione e framework (Python, R)
- Data Analytics (Alteryx, Apache Spark, SAS, Big ML, MATLAB, Jupyter)
- Data Visualisation (Tableau, RawGraphs, ChartBlocks, PowerBI, QlikSense)
- Structured data
- Unstructured data
- Dati grezzi (.pdf, .csv, .txt)

TECNICHE TIPICHE DI REALIZZAZIONE/CONDUZIONE DELLE ATTIVITÀ

- Processi semi-automatici (ontologies) di text mining
- Tecniche di data cleaning, exploration
- Natural Language Processing (NLP)
- Forecasting modeling
- Clustering modelling
- Classification modelling
- Tecniche di data visualisation
- Predictive analytics
- Prescriptive analytics
- Automated analytics

OUTPUT TIPICI DELLE ATTIVITÀ

- Standardizzazione del processo di data curation svolta
- Data curation e data quality control effettuate
- Modelli, algoritmi e strumenti per l'analisi avanzate di dataset effettuati
- Rapporto finale redatto

INDICAZIONI A SUPPORTO DELLA SCELTA DEL METODO VALUTATIVO E DELLA PREDISPOSIZIONE DELLE PROVE

ESTENSIONE SUGGERITA DI VARIETÀ PRESTAZIONALE

1. Le diverse tipologie di data curation e data quality control
2. L'insieme degli strumenti per la raccolta e l'analisi dei dati (Natural Language Processing (NLP), Forecasting modeling, Clustering modelling, Classification modelling)
3. L'insieme di tecniche e tools per un'efficace visualizzazione dei dati (data visualisation)
4. Un set di obiettivi di analisi
5. Un set di dati grezzi

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

DISEGNO TIPO DELLA VALUTAZIONE

1. Prova prestazionale: a partire da dataset, rappresentazione del processo di data curation e quality control, creazione di modelli e algoritmi ed illustrazione dei risultati mediante procedure di data visualisation
2. Colloquio tecnico relativo al processo di data curation e quality control, elaborazione e reporting mediante data visualisation per una tipologia di analisi non oggetto di prova prestazionale

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

SCHEDA DI CASO

RISULTATO ATTESO 3 - Pianificare e coordinare i processi di data analytics, assicurando il corretto feeding dei livelli strategici e decisionali

1 - PIANIFICAZIONE E COORDINAMENTO DEI PROCESSI

Grado di complessità 1

1.1 MODELLO DATA DRIVEN

Pianificare un modello organizzativo per la gestione dei dati business driven, in cui le competenze di Data Science sono localizzate all'interno delle singole unità di business

1.1 MODELLO CENTRALIZZATO

Pianificare un approccio centralizzato per la gestione dei dati, identificando una specifica figura (Chief Data Scientist) che coordini le risorse in una struttura dedicata

1.1 MODELLO "A MATRICE"

Pianificare un modello organizzativo per la gestione dei dati in cui il responsabile (Chief Data Scientist) pianifica le risorse distribuite all'interno delle aree di business

2 - TRASFERIMENTO ALLE STRATEGIE DI BUSINESS

Grado di complessità 3

2.3 COLLEGAMENTO CON LA STRATEGIA AZIENDALE

Favorire lo svolgimento di operazioni real-time tramite l'utilizzo di asset strategici, in grado di convertire istantaneamente il processo di business in atto, fornendo anche dei feedback in tempo reale

Grado di complessità 2

2.2 SUPPORTO AI PROCESSI DECISIONALI

Supportare il processo decisionale fornendo opzioni di scelta al top management, basate sull'analisi di grandi volumi di dati, così da fornire soluzioni applicabili in tempo reale

2.2 SUPPORTO ALLA CUSTOMER ORIENTATION

Supportare azioni di fidelizzazione e customer care sulla base dell'analisi di grandi volumi di dati che consentano maggiore accuratezza nella profilazione dei clienti o personalizzazione del servizio

Grado di complessità 1

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

2.1 SUPPORTO ALLO SVILUPPO DI NUOVI PRODOTTI/SERVIZI

Contribuire all'ideazione di nuovi prodotti o servizi innovativi grazie alle idee provenienti dai risultati dei processi di data analytics e business intelligence

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

SCHEDA RISORSE A SUPPORTO DELLA VALUTAZIONE DEL RISULTATO ATTESO 3

RISORSE FISICHE ED INFORMATIVE TIPICHE (IN INPUT E/O PROCESS ALLE ATTIVITÀ)

- Norme in materia di protezione della privacy e protezione dati personali (GDPR)
- Criteri generali di Data Analytics (Knowledge Discovery in Databases)
- Modelli organizzativi

TECNICHE TIPICHE DI REALIZZAZIONE/CONDUZIONE DELLE ATTIVITÀ

- Tecniche di pianificazione e coordinamento dei processi di Data Science & Analytics
- Tecniche di conduzione di analisi di business environment
- Tecniche di trasferimento del know-how data-driven a supporto delle strategie di business
- Metodi e tecniche di management di organizzazione aziendale

OUTPUT TIPICI DELLE ATTIVITÀ

- Analisi del business environment condotta
- Unità di business data-driven pianificate
- Chief Data Scientist e struttura per la gestione dei dati identificate
- Chief Data Scientist ed aree di business data-driven identificate
- Nuovi prodotti e servizi data-driven offerti
- Customer-orientation effettuata
- Supporto data-driven ai processi di decision-making dell'organizzazione realizzato

INDICAZIONI A SUPPORTO DELLA SCELTA DEL METODO VALUTATIVO E DELLA PREDISPOSIZIONE DELLE PROVE

ESTENSIONE SUGGERITA DI VARIETÀ PRESTAZIONALE

1. L'insieme dei fondamenti Data Analytics (Knowledge Discovery in Databases) applicati al business
2. Le diverse tipologie di governance dei processi di data analytics (modello centralizzato, modello data-driven, Modello a matrice)
3. Un set di casi (requisiti normativi, best practicesm posizionamento aziendale)

DISEGNO TIPO DELLA VALUTAZIONE

1. Prova prestazionale: sulla base del set dato, impostazione della strategia di governance dei dati
2. Colloquio tecnico relativo ai fondamenti di Data Analytics e di management applicati al business non oggetto di prova prestazionale

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

FONTI

Cesarini, M., Fontana, M. Mercurio, F., Mezzanica M. & Vegetti, N., Data Quality: Un Approccio Metodologico e Applicativo. “Il caso delle COB del mercato del lavoro in Lombardia”, Working Paper “Documentazione relativa alla ricostruzione di una metodologia unificata, ripetibile e aperta, CRISP E ARIFL EMC (2015), Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, John Wiley & Sons, Indianapolis

European e-Competence Framework (E-CF), v.04: <https://www.ecompetences.eu/>

Freitas, A. & Curry E. (2016). Big Data Curation, In Cavanillas, J.M & Curry E. (2016). New Horizons for a Data-Driven Economy. A Roadmap for Usage and Exploitation of Big Data in Europe, Big Data Public Private Forum, Springer Open, p. 87-119

Gelernter, J. & Lesk, M. (2011), Use of Ontologies for Data Integration and Curation. In The International Journal of Digital Curation, Issue 1, Volume 6

SITOGRAFIA

MediaSapens, Data Curation: the essential step for integrated data-driven research