

Proposte di Attività per Tirocini Aziendali

Proposta 1

Titolo dell'Attività:

Progettazione, implementazione e validazione sperimentale di applicazioni per l'analisi dei dati provenienti da scenari *Internet of Things* (IoT)

Descrizione dell'Attività:

I dispositivi IoT generano un'enorme quantità di dati che può essere adoperata per monitorare lo stato e le attività degli oggetti o asset che costituiscono l'ambiente circostante. L'attività è focalizzata principalmente sulla realizzazione di un'applicazione prototipo in grado di eseguire analisi "intelligenti" real-time e batch sui dati provenienti da tali dispositivi con l'obiettivo primario di discernere le informazioni salienti e dedurne nuove.

Nello specifico, il tirocinante lavorerà, insieme ad un gruppo di persone, alla realizzazione di un prototipo contraddistinto dalle seguenti caratteristiche: analisi dei dati storici e real-time per eseguire predizioni sul comportamento futuro o su eventi sconosciuti e rari; analisi descrittiva e predittiva (e.g., manutenzione preventiva dei guasti); identificazione dei fattori che positivamente e negativamente influenzano lo stato di salute e le performance di un oggetto reale; modifica dello stato dei dispositivi per intraprendere azioni necessarie al fine di raggiungere determinati risultati o obiettivi; agevolazione dei processi decisionali.

Una parte dell'attività sarà dedicata alla validazione sperimentale dei risultati con enfasi sull'efficacia e l'efficienza delle analisi previste.

Dettaglio attività oggetto della proposta:

- Analisi stato dell'arte
- Definizioni degli obiettivi
- Selezione dei materiali, metodi e tecniche
- Progettazione applicazione
- Realizzazione prototipo
- Validazione sperimentale dei risultati e verifica della congruità con gli obiettivi prefissati

Le attività prevedranno anche verifiche periodiche e discussioni con l'Area R&D dell'azienda.

Tematiche Coinvolte:

- Internet of Things
- Filtraggio, trasformazione, normalizzazione e integrazione di eventi/allarmi
- Big Data Analytics
- Real time e batch analysis
- Analisi "intelligente" dei dati
- Data mining
- *Complex Event Processing* (CEP)
- Prescriptive Maintenance e Predictive Maintenance

- **Statistica di base**
 - Distribuzioni di frequenze, tabelle e grafici
 - Valori medi
 - Variabilità
 - Dipendenza e correlazione
- **Machine Learning**
 - Clustering
 - Classificazione
 - Regressione
 - Reti Neurali
 - *Recurrent Neural Network (RNN)* basate su *Long Short-Term Memory (LSTM)*
- Altro:
 - Sistemi paralleli e distribuiti
 - Parallelismo implicito
 - Reti di calcolatori
 - Cloud e Grid Computing
 - File system e database distribuiti
 - Sicurezza e privacy
 - Sistemi embedded
 - Oggetti intelligenti

Principali Risultati Previsti:

- Prototipo dell'applicazione in grado di analizzare in maniera "intelligente" una grossa mole di dati
- Validazione sperimentale dell'efficacia e dell'efficienza delle tecniche adottate.

Conoscenze e Competenze Richieste agli Allievi:

- Algoritmi e librerie di Machine & Deep Learning
- Competenze su sistemi distribuiti
- Programmazione Java e/o Python
- Conoscenza (anche preliminare) del framework Apache Spark.

Proposta 2

Titolo dell'Attività:

Studio, implementazione e validazione piattaforma Kubernetes con gestione dinamica del carico di lavoro.

Contesto dell'area di ricerca oggetto di attività

Il contesto dell'area di svolgimento del tirocinio rientra nelle attività aziendali di ricerca e sviluppo orientate allo studio, all'implementazione e alla validazione delle prassi e delle tecnologie per l'esercizio in ambiente enterprise di strumenti e sistemi per l'orchestrazione di microservizi in private cloud, attraverso l'impiego delle piattaforme OpenStack e Kubernetes.

Descrizione dell'Attività:

L'obiettivo del percorso è la progettazione e installazione di un'infrastruttura Kubernetes composta da un sistema centrale e diversi elementi periferici (edge) con successiva implementazione delle logiche dinamiche adottate per la distribuzione del carico di lavoro in base alle policies implementate.

Nel corso delle attività, il tirocinante coadiuverà il team aziendale nello studio, nella progettazione, nell'ingegnerizzazione, nel testing e nella validazione della piattaforma complessiva con monitoraggio dei carichi di lavoro eseguiti.

Durante le attività saranno adoperati strumenti ed API fornite da Kubernetes per la gestione del ciclo di vita delle applicazioni, monitoraggio e accesso ai dati generati e organizzazione complessiva della piattaforma, con particolare riferimento al sistema di scheduling basato su policies.

Le attività del laureando saranno divise in varie fasi:

- Analisi dello stato dell'arte e della letteratura oggetto del tirocinio
- Selezione dei metodi e materiali
- Installazione e configurazione piattaforma Kubernetes
- Installazione e configurazione sistemi periferici
- Validazione sperimentale e testing
- Disseminazione del lavoro e dei risultati.

Le attività prevedranno anche verifiche periodiche e discussioni con l'Area R&D dell'azienda.

Tematiche Coinvolte:

- Docker e sistemi basati su container
- Software defined Networking
- Architetture cloud
- Ambiente Kubernetes
- Applicazioni distribuite
- Bilanciamento dinamico dei carichi di lavoro

Principali Risultati Previsti:

- Prototipo ambiente distribuito Kubernetes
- Validazione sperimentale dell'efficacia e dell'efficienza delle tecniche adottate.

Conoscenze e Competenze Richieste agli Allievi:

- Networking di base
- Competenze su sistemi distribuiti
- Sistemi Linux
- Nozioni di infrastrutture di server, rete e storage;
- Attitudine allo studio e al lavoro di gruppo
- Programmazione di base

Proposta 3

Titolo dell'Attività:

Studio, implementazione e test front-end per editing visuale di schemi relativi a infrastrutture ICT.

Contesto dell'area di ricerca oggetto di attività

Il contesto dell'area di svolgimento del tirocinio rientra nelle attività aziendali legate allo sviluppo evolutivo di una piattaforma cyber range finalizzata all'addestramento e formazione di personale tecnico. In tale ambito System Management intende sviluppare un sistema per la composizione visuale degli ambienti di gioco mediante la realizzazione di una web application avanzata.

Descrizione dell'Attività:

L'obiettivo del percorso è la progettazione e realizzazione prototipale di una moderna applicazione web basata su tecnologia Angular dotata della capacità di costruire schemi e diagrammi di infrastrutture ICT in modo visuale mediante drag&drop di elementi e interconnessione degli stessi.

Nel corso delle attività, il tirocinante coadiuverà il team aziendale nello studio, nella progettazione, nell'ingegnerizzazione, nel testing e nella validazione dello strumento sperimentale con produzione di alcuni scenari ICT.

Durante le attività saranno adoperati strumenti ed API implementati da System Management per la propria piattaforma Cyber Range con particolare riferimento al modello di rappresentazione degli scenari di gioco.

Le attività del laureando saranno divise in varie fasi:

- Analisi dello stato dell'arte e della letteratura oggetto del tirocinio
- Studio modello rappresentazione ambienti di gioco, standard TOSCA
- Sistemi basati su paradigma Infrastructure-as-a-Code
- Progettazione applicazione
- Realizzazione prototipo
- Validazione sperimentale dei risultati e verifica della congruità con gli obiettivi prefissati

Le attività prevedranno anche verifiche periodiche e discussioni con l'Area R&D dell'azienda e con il gruppo responsabile del prodotto Cyber Range.

Tematiche Coinvolte:

- Ambienti web Front-end
- Single page web application e Progressive web application
- Framework Angular
- Linguaggio TOSCA e approccio Infrastructure-as-a-Code

Principali Risultati Previsti:

- Prototipo strumento per visual design di ambienti ICT di addestramento
- Validazione sperimentale dell'efficacia e dell'efficienza delle tecniche adottate.

Conoscenze e Competenze Richieste agli Allievi:

- Programmazione di base

- Nozioni su web front-end, CSS, HTML
- Gradita conoscenza di base framework Angular
- Nozioni di base su JSON e API HTTP/REST
- Nozioni di infrastrutture di server, rete e storage;
- Attitudine allo studio e al lavoro di gruppo

Proposta 4

Titolo dell'Attività:

Sviluppo e validazione sperimentale di applicazioni per l'analisi dei dati sfruttando tecniche di Machine e Deep Learning per l'Anomaly Detection

Descrizione dell'Attività:

L'attività oggetto di tirocinio e tesi riguarda l'analisi di grosse moli di dati provenienti da eventi scaturiti da sensori e dispositivi *Internet of Things* (IoT). L'obiettivo principale è realizzare un motore di analisi che dinamicamente possa estrarre semantica dagli eventi con lo scopo di determinare modelli comportamentali dei dati. Per esempio, saranno oggetto di studio: l'identificazione della distribuzione statistica dei dati in condizioni "normali"; l'identificazione di template e strutture dei dati; la rilevazione automatica di outlier; la creazione e l'adattamento dinamico di regole al fine di evidenziare comportamenti anomali rispetto alla baseline; l'adozione di modelli predittivi per anticipare fenomeni ed eventi usando le serie storiche; analisi dei dati storici e real-time per eseguire predizioni sul comportamento futuro o su eventi sconosciuti e rari; analisi descrittiva e predittiva (e.g., manutenzione preventiva dei guasti); apprendere i comportamenti degli utenti nell'ambito delle attività di ricerca eventi e sottoscrizioni di notifiche; etc.

Oltre allo studio e all'analisi dei metodi presenti in letteratura, sarà prevista un'attività di ricerca delle librerie e dei framework che consentano di adoperare e/o implementare tali algoritmi. Successivamente sarà prevista una fase di validazione sperimentale che confronti una o più tecniche in base a fissati *Key Performance Indicator* (KPI) adoperando un unico dataset come testbed.

Dettaglio attività oggetto della proposta:

- Analisi stato dell'arte
- Definizioni degli obiettivi
- Selezione dei materiali, metodi e tecniche
- Progettazione applicazione
- Realizzazione prototipo
- Validazione sperimentale dei risultati e verifica della congruità con gli obiettivi prefissati.

Le attività prevedranno anche verifiche periodiche e discussioni con l'Area R&D dell'azienda.

Tematiche Coinvolte:

- Internet of Things
- Filtraggio, trasformazione, normalizzazione e integrazione di eventi/allarmi
- Big Data Analytics
- Real time e batch analysis
- Analisi "intelligente" dei dati
- Data mining
- *Complex Event Processing* (CEP)
- Prescriptive Maintenance e Predictive Maintenance
- **Statistica di base**
 - Distribuzioni di frequenze, tabelle e grafici
 - Valori medi

- Variabilità
- Dipendenza e correlazione
- **Machine & Deep Learning**
 - Unsupervised Learning
 - Reinforced Learning
 - Clustering
 - Classificazione
 - Regressione
 - Reti Neurali
 - *Recurrent Neural Network (RNN)*
 - *Convolutional Neural Network (CNN)*
 - Decision Tree
- Altro:
 - Sistemi paralleli e distribuiti
 - Parallelismo implicito
 - Reti di calcolatori
 - Cloud e Grid Computing
 - File system e database distribuiti
 - Sicurezza e privacy
 - Sistemi embedded
 - Oggetti intelligenti

Principali Risultati Previsti:

- Prototipo dell'applicazione in grado di analizzare in maniera "intelligente" una grossa mole di dati usando sistemi di Machine/Deep Learning
- Validazione sperimentale dell'efficacia e dell'efficienza delle tecniche adottate.

Conoscenze e Competenze Richieste agli Allievi:

- Algoritmi e librerie di Machine Learning
- Competenze su sistemi distribuiti
- Programmazione Java/Scala/Python/R
- Conoscenza (anche preliminare) del framework Apache Spark
- Attitudine allo studio e al lavoro di gruppo.

Proposta 5

Titolo dell'Attività:

Analisi, prototipizzazione e validazione sperimentale di algoritmi per l'identificazione di anomalie e la predizione di eventi su serie temporali

Contesto dell'area di ricerca oggetto di attività:

In letteratura sono disponibili diversi approcci per la rilevazione di anomalie su serie temporali sia attraverso processi di *training* che adoperando *self-training* del sistema. L'obiettivo principale di questi approcci è isolare un modello sperimentale di analisi dati basato su tecniche Machine & Deep Learning (anche combinandole con l'utilizzo di algoritmi statistici di base) per rilevare comportamenti "normali" e "anomali" dei dati. Individuare tali deviazioni rispetto al comportamento atteso all'interno di dati temporali rappresenta un aspetto importante per assicurare il normale comportamento dei sistemi oppure evidenziare in modo repentino variazioni critiche. Nel dominio applicativo di queste tecniche è importante modellare cosa si ritiene "anomalo" (e quanto anomala è la deviazione rilevata) rispetto a cosa rappresenta un comportamento "normale", osservandone il comportamento di un sistema nel tempo, rilevandone e anticipandone pattern periodici e stagionali, e correlandone ad altri sistemi e/o fenomeni.

In particolare, in questo contesto rientrano le seguenti attività: l'identificazione della distribuzione statistica dei dati in condizioni "normali"; l'identificazione di template e strutture dei dati; la rilevazione automatica di outlier; la creazione e l'adattamento dinamico di regole al fine di evidenziare comportamenti anomali rispetto alla baseline; l'adozione di modelli predittivi per anticipare fenomeni ed eventi usando le serie storiche; analisi dei dati storici e real-time per eseguire predizioni sul comportamento futuro o su eventi sconosciuti e rari; analisi descrittiva e predittiva (e.g., manutenzione preventiva dei guasti).

Descrizione dell'Attività:

Le attività di tirocinio, che rientrano nel contesto presentato in precedenza, prevedranno, in prima battuta, lo studio e successivamente la selezione di algoritmi per l'analisi delle serie temporali al fine di rilevare anomalie e/o anticiparle. Ad esempio, oggetto di analisi saranno le tecniche di *Univariate / Multivariate Anomaly Detection*, oltre che l'analisi e la predizione di trend e stagionalità nelle serie temporali. In questo scenario, gli studi prevedranno anche l'analisi di tecnologie basate su reti neurali profonde, ad esempio, *Recurrent Neural Networks (RNN)* di tipo *Long Short-Term Memory Networks (LSTM)* ottimizzate per operare su serie temporali di dati (sia in real-time sia long-term), e classificare le informazioni raccolte con rilevamento di anomalie. Inoltre, saranno oggetto di indagine anche diversi modelli statistici, come *Autoregressive Integrated Moving Average (ARIMA)* e *Autoregressive Moving Average Model (ARMA)*, per la comprensione e la previsione dei valori delle caratteristiche nelle serie temporali. E' da precisare che gli specifici algoritmi e tecniche da sperimentare, saranno selezionati al termine di una prima analisi dello stato dell'arte.

Nel corso delle attività di sperimentazione saranno utilizzati anche strumenti di prototipazione rapida quali, ad esempio, PySpark, TensorFlow, Turi e/o librerie DL basate su Python, per costruire modelli sperimentali su cui svolgere le attività di verifica sui comportamenti e performance degli algoritmi individuati.

Il laureando disporrà di dataset di benchmarking, sia reali e privati che simulati, per la validazione dei selezionati algoritmi di analisi dati. Infatti, un ruolo predominante verrà dato alla fase di confronto e validazione sperimentale degli individuati modelli di analisi su tali dataset. In particolare, saranno adoperati

opportuni *Key Performance Indicator* (KPI) per analizzare sia l'efficacia che l'efficienza dei modelli (*Experimental Analysis of Algorithms*).

Le attività del laureando saranno divise in varie fasi:

- Analisi dello stato dell'arte e della letteratura oggetto del tirocinio
- Selezione dei metodi e materiali (e.g., algoritmi, dataset, KPI, etc.)
- Implementazione con strumenti di prototipizzazione rapida (e.g., PySpark)
- Validazione sperimentale con KPI
- Disseminazione del lavoro e dei risultati.

Le attività prevedranno anche verifiche periodiche e discussioni con l'Area R&D dell'azienda.

Tematiche Coinvolte:

- Internet of Things
- Filtraggio, trasformazione, normalizzazione e integrazione di eventi/allarmi
- Big Data Analytics
- Real time e batch analysis
- *Complex Event Processing* (CEP)
- *Artificial Intelligence for IT Operations* (AIOps)
- Outlier detection
- Anomaly detection
- Prescriptive Maintenance, Predictive & Preemptive Maintenance
- Statistica di base
 - Distribuzioni di frequenze, tabelle e grafici
 - Valori medi
 - Variabilità
 - Dipendenza e correlazione
- Machine & Deep Learning
 - Supervised Learning
 - Unsupervised Learning
 - Reinforced Learning
 - Continuous Learning
 - Clustering
 - Classificazione
 - Regressione
 - Reti Neurali
 - *Recurrent Neural Network* (RNN)
 - *Convolutional Neural Network* (CNN)
 - Decision Tree
- Altro
 - Sistemi paralleli e distribuiti
 - Parallelismo implicito
 - Cloud e Grid Computing
 - File system e database distribuiti.

Principali Risultati Previsti:

- Prototipo dell'applicazione
- Validazione sperimentale dell'efficacia e dell'efficienza delle tecniche adottate.

Conoscenze e Competenze Richieste agli Allievi:

- Algoritmi e librerie di Machine & Deep Learning
- Competenze su sistemi distribuiti
- Programmazione Java/Scala/Python/R
- Conoscenza (anche preliminare) del framework Apache Spark
- Attitudine allo studio e al lavoro di gruppo

Proposta 6

Titolo dell'Attività:

Studio, implementazione e validazione architetture e servizi per la raccolta, l'analisi, la memorizzazione e la ricerca di dati provenienti da sistemi IoT

Contesto dell'area di ricerca oggetto di attività

Il contesto dell'area di svolgimento del tirocinio rientra nelle attività aziendali di ricerca e sviluppo orientate allo studio, all'implementazione e alla validazione di un sistema a supporto della raccolta, dell'analisi, della memorizzazione e della ricerca di dati generati da dispositivi e sistemi IoT. In particolare, rientra in tale area di lavoro sia la gestione della persistenza di grandi quantità di dati adottando diverse strategie di archiviazione finalizzate ad ottimizzarne il successivo utilizzo.

Vengono adoperati sia linguaggi di programmazione con librerie per la gestione dei Big Data (e.g., Apache Spark con applicazioni Java o Python, etc.) e motori di archiviazione journal-based, time-series, colonnari e document oriented.

Descrizione dell'Attività:

All'interno del contesto di ricerca e sviluppo, presentato in precedenza, il lavoro del tirocinante riguarderà lo studio, l'implementazione e la validazione dei processi di analisi e interrogazione sui dati strutturati per diverse finalità di utilizzo all'interno di una piattaforma IoT. Nel corso delle attività, il tirocinante coadiuverà il team aziendale nello studio, nella progettazione, nell'ingegnerizzazione, nel testing e nella validazione sperimentale di algoritmi per analisi avanzata dei dati operando su elementi grezzi (raw data), serie temporali e insiemi pre-aggregati.

In particolare, durante le attività saranno adoperati strumenti ed API per la ricerca dei dati in grado di eseguire query spazio-temporali. Inoltre, saranno oggetto di lavoro metodi e tecniche per condividere, elaborare e presentare dati. Il candidato si occuperà anche di integrare un sistema di dashboard avanzato con lo scopo di effettuare presentazioni dei dati a supporto delle decisioni.

Le attività del laureando saranno divise in varie fasi:

- Analisi dello stato dell'arte e della letteratura oggetto del tirocinio
- Selezione dei metodi e materiali
- Attività di sviluppo sperimentale
- Attività di validazione sperimentale e testing
- Disseminazione del lavoro e dei risultati.

Le attività prevedranno anche verifiche periodiche e discussioni con l'Area R&D dell'azienda.

Tematiche Coinvolte:

- Big Data Analysis
- IoT System
- Sistemi database e strategie di archiviazione avanzate
- Query su time-series, query su raw data e query spazio-temporali

Principali Risultati Previsti:

- Prototipo dell'applicazione
- Validazione sperimentale dell'efficacia e dell'efficienza delle tecniche adottate.

Conoscenze e Competenze Richieste agli Allievi:

- Competenze su sistemi distribuiti
- Programmazione Java e Python
- Conoscenza (anche preliminare) del framework Apache Spark
- Attitudine allo studio e al lavoro di gruppo