



Simone Cipriani <sim.cipr@gmail.com>

Re: Progetto per Tirocinio Interno

Donatella Merlini <donatella.merlini@unifi.it>

5 marzo 2018 09:31

A: Simone Cipriani <sim.cipr@gmail.com>

Sì, nel mio ufficio alle 10.30

A lunedì prossimo
Donatella Merlini

Il 3/4/2018 4:50 PM, Simone Cipriani ha scritto:
Va benissimo. Nel suo ufficio?

Saluti,
Simone

Il giorno 04 mar 2018, alle ore 15:50, Donatella Merlini <donatella.merlini@unifi.it> ha scritto:

Verso le 10.30 va bene oppure preferisce più tardi?

Donatella Merlini

Il Sabato 03/03/2018 16:08 Simone Cipriani ha scritto:

Preferirei in mattinata, almeno torno per pranzo in ufficio e faccio il pomeriggio.
Non si preoccupi, non mi dimenticherò Informatica Teorica e Calcolo Numerico (anche se non sa quanto mi piacerebbe poterlo fare).

Buon fine settimana,
Simone

Il giorno 3 marzo 2018 11:09, Donatella Merlini <donatella.merlini@unifi.it> ha scritto:

La settimana prossima sono messa piuttosto male, se ci vedessimo lunedì 12 marzo per lei può andare bene? In tarda mattinata o nel primo pomeriggio, come preferisce. Tra l'altro dobbiamo anche attivare il tirocinio in modo formale. Mi raccomando però, dia priorità agli esami che le mancano.

Mi faccia sapere, a presto

Donatella Merlini

Il Venerdì 02/03/2018 21:20 Simone Cipriani ha scritto:

Ho capito. La produttività del 2014 semmai si potrebbe aggiungere in seguito, alla fine aprile è tra un mese.
Per vedersi le andrebbe bene verso la fine della prossima settimana? Così ho un po' di tempo per prendere familiarità con i dati.
Saluti,
Simone

Il giorno 2 marzo 2018 17:56, Donatella Merlini <donatella.merlini@unifi.it>

<mailto:donatella.merlini@unifi.it>> ha scritto:

Intanto le allego i file con le valutazioni relative agli anni accademici di riferimento. Dato che le valutazioni del CdS sono visibili a tutti non mi sono preoccupata di eliminare i riferimenti ai docenti. Ovviamente però nella tesi non dovrà esserci riferimento esplicito ai docenti ma solo ai corsi ed eventualmente ai cambi di docente. La scheda di valutazione negli anni è cambiata e non escludo che ci siano delle differenze in quelle che le ho mandato. Comunque dai file .csv dovrebbe essere possibile risalire all'organizzazione della scheda in ogni anno. Per quanto riguarda i dati sulle carriere degli studenti, per il momento non sono in grado di mandarle niente di nuovo perché per la coorte 2014 le dovrei dare la produttività fino ad aprile 2018 che ancora non è disponibile. Quindi almeno per il momento direi di considerare gli stessi dati che le avevo mandato per il progetto di DMO. Se necessario glieli rimando.

Poi magari fra qualche giorno ci vediamo per parlarne di persona.

Buon lavoro
Donatella Merlini

Il 3/1/2018 4:23 PM, Simone Cipriani ha scritto:

Ottimo, aspetto sue notizie allora.

Saluti,
Simone

Il giorno 1 marzo 2018 14:49, Donatella Merlini
<donatella.merlini@unifi.it>
<mailto:donatella.merlini@unifi.it>
<mailto:donatella.merlini@unifi.it>
<mailto:donatella.merlini@unifi.it>> ha scritto:

Secondo me riusciamo a mettere su un progetto interessante. Se vuole usare MongoDB, anche se forse non è strettamente necessario, per me va bene. Penso anche che Weka come strumento di analisi sia adeguato.

Mi dia qualche giorno di tempo per procurarle i dati, forse riesco a darle quelli degli studenti aggiornati e poi le devo dare i dati delle valutazioni. Magari poi ci vediamo per fare il punto della situazione e organizzare il lavoro.

A presto

Donatella Merlini

Il Mercoledì 28/02/2018 20:45 Simone Cipriani ha scritto:

Buongiorno prof.ssa Merlini,
ho pensato un po' alla faccenda e mi sono fatto un'idea di quello che si potrebbe fare. Le condivido le mie considerazioni, così mi può dire se ho ragionato nel modo giusto.

Nel merito dei dati, potrebbe essere interessante studiare i dataset che ha proposto lei, magari mettendo i corsi al centro dell'attenzione e cercando ad esempio correlazioni fra i risultati e valutazione della didattica, oppure fra vari esami connessi dal vincolo di propedeuticità, o altri aspetti che possono venire in mente avendo i dati alla mano. Insomma, c'è del potenziale per trovare informazioni utili.

Per come lo sto immaginando io, il lavoro da fare sarebbe:

1. progettare ed implementare uno schema che descrive tutti i dati a disposizione
2. lanciare algoritmi di data mining su sottoinsiemi di dati preprocessati appositamente

Quindi, la nostra stack sarebbe composta da un dbms ed un software tipo Weka.

Per il db, i nostri dati sono statici e le nostre query saranno in pratica restituzioni di viste, quindi i vantaggi principali del paradigma NoSQL che sono la scalabilità, l'estensibilità e le performance non ci interessano. Può essere utile però la flessibilità del modello a documenti di MongoDB per esprimere le relazioni fra i nostri due dataset di natura diversa; niente che non si possa fare nel modello relazionale, ma potrebbe essere più facile. A parità di vantaggi io sarei tendente a scegliere MongoDB, però il mio è un bias da attrazione verso la "next-big-thing", dovrei pensarci meglio con i dati alla mano.

Per il data mining, Weka è una scelta solida (gira su Java e si può connettere a tutti i dbms per cui ci siano i jdbc, inoltre le sue API sono ben documentate se si volesse estendere o modificare gli algoritmi già forniti). Comunque esistono decine di altre

soluzioni, ma dovrei avere un'idea più chiara
di cosa si vuole
minare prima di valutare eventuali alternative.
Lei che ne pensa?
Mi scuso di nuovo per il wall.of-text, ma
ormai la mia prolissità
non è più un mistero.

Saluti,

Simone

Il giorno 25 febbraio 2018 18:48, Donatella
Merlini

<donatella.merlini@unifi.it
<mailto:donatella.merlini@unifi.it>
<mailto:donatella.merlini@unifi.it
<mailto:donatella.merlini@unifi.it>>>
ha scritto:

Ho letto la mail fino in fondo ;-)

Si', certo, sono disponibile a farle da
relatore, il problema
e' capire cosa posso proporle. Non mi
stupisce che i referenti
della sua azienda le abbiano detto che
non possano darle i
dati, ma un tentativo era giusto farlo.

Per quanto riguarda MongoDB, ho seguito
una tesi
sull'argomento generale 1-2 anni fa e in
questo momento c'e'
un secondo studente che sta svolgendo una
tesi che consiste
nel realizzare una base di dati sia in
MongoDB che con un
classico sistema relazionale, con
l'obiettivo di fare un
confronto. Ovviamente potrebbe esserci
ancora spazio in questa
direzione, magari con un progetto in cui
i dati sono al centro
dell'attenzione, ad esempio, di nuovo
quelli relativi agli
studenti di informatica. Si potrebbe
pensare di studiare
insieme a questi anche altri tipi di
dati: ad esempio, quelli
relativi alla valutazione della didattica
negli anni
accademici di riferimento, anche se in
questo caso si
avrebbero a disposizione solo dati in
forma aggregata e non a
livello di valutazione data del singolo
[studente](#) e quindi
queste informazioni non potrebbero essere
messe in relazione
in modo esatto.

Non so se in un contesto di questo tipo
l'uso di MongoDB possa
portare a qualche vantaggio, oppure se

sia preferibile
lavorare direttamente nel modello
relazionale. Pensiamoci
qualche giorno, se ha delle proposte
diverse ovviamente sono
benvenute.

Cordiali saluti

Donatella Merlini

Il Sabato 24/02/2018 15:16 Simone
Cipriani ha scritto:

Buongiorno prof.ssa Merlini,
intanto le chiedo scusa, le avevo
detto alla fine
dell'esame di Data Mining che mi
sarei fatto sentire nei
giorni seguenti, ma ho deciso di
provare a dare
Informatica Teorica nell'appello di
febbraio (ovviamente,
senza successo) ed i giorni sono
diventati settimane.
Le volevo fare diverse domande, le
scrivo qui sotto, ma se
non volesse perdere tempo a leggere
il mio wall-of-text ma
preferisse parlarne a voce mi dica e
mi organizzo per
incontrarci.

Intanto, è sempre disposta a farmi da
relatrice?

Nel merito del lavoro da fare, ho
chiesto ai referenti di
BHGE con i quali ho un po' di
confidenza se è possibile
utilizzare i dati dei compressori
centrifughi sui quali
lavoriamo per fare qualche studio, ma
sono (siamo) tutti
sotto uno strettissimo NDA e nessuno
può autorizzare una
divulgazione del genere. Potrei
chiedere a qualche
"gerarca" più in alto in BHGE, ma
preferirei non farlo,
alla fine rappresento un piccolo
fornitore che lavora per
loro e potrebbero non prenderla bene.
Potremmo pensare a qualcos'altro?

Stavo giusto in questi
giorni seguendo un progetto per un
altro cliente nel quale
stiamo cercando di usare MongoDB (un
dbms NoSQL) per
mantenere dei dati e mi sembra una
tecnologia affascinante
che ha un sacco di applicazioni sui
Big Data. Potrebbe
interessarle un argomento del genere?

Se no proponga pure
altre idee, sono aperto a quasi tutto.
Inoltre, volevo chiederle anche un
consiglio su un altro
piano: sarebbe opportuno secondo lei
cominciare il lavoro
per il tirocinio adesso?
Secondo me, sì, vista la mia situazione:
- mi mancano IT e CN che prevedo di
dare entrambi negli
appelli estivi.
- vorrei laurearmi il prima
possibile, luglio o, più
realisticamente, ottobre
Facendo il lavoro per il tirocinio
nei prossimi mesi,
potrei dare i due esami che mi
mancano nella sessione
estiva (o a settembre al massimo) e
finalmente laurearmi.
Sbaglio qualcosa in questo ragionamento?
Se è arrivata a leggere fin qui, la
ringrazio
infinitamente per la sua pazienza e
per il tempo che mi ha
dedicato.
Buon fine settimana,
Simone

--

=====

Donatella Merlini
Dipartimento di Statistica, Informatica, Applicazioni "G.
Parenti"
Viale Morgagni 65, 50134 Firenze
<[https://maps.google.com/?q=Viale+Morgagni+65,+50134+Firenze&entry=gmail&
source=g](https://maps.google.com/?q=Viale+Morgagni+65,+50134+Firenze&entry=gmail&source=g)>
tel. +39 055 2751509 <tel:%2B39%20055%202751509>
donatella.merlini@unifi.it
<<mailto:donatella.merlini@unifi.it>>
<http://local.disia.unifi.it/merlini/>
<<http://local.disia.unifi.it/merlini/>>

=====

--

=====

Donatella Merlini
Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti"
Viale Morgagni 65, 50134 Firenze
tel. +39 055 2751509
donatella.merlini@unifi.it
<http://local.disia.unifi.it/merlini/>

=====