UNIVERSITÀ DEGLI STUDI DI MILANO

SELEZIONE PUBBLICA, PER TITOLI ED ESAMI PER IL RECLUTAMENTO DI N. 1 UNITÀ DI TECNOLOGO DI SECONDO LIVELLO CON RAPPORTO DI LAVORO SUBORDINATO A TEMPO DETERMINATO DELLA DURATA DI 28 MESI PRESSO L'UNIVERSITÀ DEGLI STUDI DI MILANO - DIPARTIMENTO DI ECONOMIA, MANAGEMENT E METODI QUANTITATIVI BANDITA CON DETERMINA N. 8825 DEL 27.5.2021, PUBBLICATA SUL SITO INTERNET DELL'ATENEO IN DATA 27.5.2021 – CODICE 21611

La Commissione Giudicatrice del concorso, nominata con determina n. 10203 del 16.6.2021, composta da:

Prof.ssa Silvia Salini - Presidente

Prof.ssa Alessandra Micheletti - Componente

Dott.ssa Daniela Bagnati - Componente

Sig.ra Luisa Castellano - Segretaria

comunica i quesiti relativi alla prova orale:

Quesiti busta 1

1. Il/la candidato/a illustri le caratteristiche principali della policy di ateneo sulla gestione dei dati della ricerca

2. Si supponga che un docente dell'Università di Milano debba gestire ed analizzare dei dati, nell'ambito di un progetto quinquennale, con le seguenti caratteristiche:
    a. I dati sono costituiti da immagini satellitari, suddivise al momento in 1000 file diversi, tutte in formato jpg e in totale occupano circa 5TB di memoria
    b. Le immagini sono pubbliche e non contengono dati sensibili
    c. Il docente continuerà ad acquisire immagini nel corso del progetto, che andranno ad aggiungersi al database
    d. Le analisi che il docente deve fare sfruttano un software commerciale, per il quale il docente possiede una licenza per sistemi operativi windows
    e. Le immagini vengono analizzate a gruppi di 10, e una volta analizzate vanno conservate, almeno fino al termine del progetto, ma di norma il docente non ha più bisogno di accedervi

    Il docente non ha fondi disponibili per acquistare hardware, mentre ha fondi per acquistare servizi di calcolo e di storage.
    Quale architettura di calcolo gli consiglierebbe di utilizzare? Si scelga tra attrezzature per HPC disponibili in ateneo (piattaforma Indaco), attrezzature disponibili presso centri di calcolo esterni (ad esempio Cineca), o servizi di cloud computing esterni all'ateneo, motivando la risposta.

3. Quali ritiene che siano le competenze necessarie ed i compiti di un Data Manager all'interno del Dipartimento di Economia, Management e Metodi Quantitativi, e quale ritiene che siano le origini principali dei dati utilizzati dai membri del dipartimento?

4. Il/la candidato/a legga e traduca il seguente brano:
    In principal component analysis, we seek to maximize the variance of a linear combination of the variables. For example, we might want to rank students on the basis of their scores on achievement tests in English, mathematics, reading, and so on. An average score would provide a single scale on

which to compare the students, but with unequal weights we can spread the students out further on the scale and obtain a better ranking.
Essentially, principal component analysis is a one-sample technique applied to data with no groupings among the observations and no partitioning of the variables into subsets y and x. All the linear combinations that we have considered previously were related to other variables or to the data structure. In regression, we have linear combinations of the independent variables that best predict the dependent variable(s); in canonical correlation, we have linear combinations of a subset of variables that maximally correlate with linear combinations of another subset of variables; and discriminant analysis involves linear combinations that maximally separate groups of observations. Principal components, on the other hand, are concerned only with the core structure of a single sample of observations on p variables. None of the variables is designated as dependent, and no grouping of observations is assumed.

Quesiti busta 2

1. Il/la candidato/a illustri le caratteristiche principali di un Data Management Plan nell'ambito di progetti nazionali o europei

2. Si supponga che un docente dell'Università di Milano debba gestire ed analizzare dei dati di tipo socio-economico, nell'ambito di un progetto, con le seguenti caratteristiche:

    a. I dati sono suddivisi in 1000 file diversi, tutti in formato .csv, e in totale occupano circa 5GB di memoria
    b. I dati provengono in parte da survey raccolti dal ricercatore, che contengono dati sensibili, per i quali il ricercatore ha raccolto, dai partecipanti al survey, dei consensi informati e un'autorizzazione al trattamento dei dati, al solo scopo della ricerca collegata al progetto. Il ricercatore non può trasferire i dati a terzi e i dati andranno distrutti al termine del progetto.
    c. Le analisi che il docente deve fare includono dei permutation test, con migliaia di iterazioni durante le quali viene calcolata la stessa quantità, ma ogni volta su un sottoinsieme casuale e diverso dei 1000 file di dati. I codici sono scritti in R.

    Il docente vorrebbe ridurre i tempi computazionali per le sue analisi, ma non ha fondi disponibili per acquistare hardware, mentre ha fondi per acquistare servizi di calcolo.
    Quale architettura di calcolo gli consiglierebbe di utilizzare? Si scelga tra attrezzature per HPC disponibili in ateneo (piattaforma Indaco), attrezzature disponibili presso centri di calcolo esterni (ad esempio Cineca), o servizi di cloud computing esterni all'ateneo, motivando la risposta.

3. Quali ritiene che siano le competenze necessarie ed i compiti di un Data Manager all'interno del Dipartimento di Economia, Management e Metodi Quantitativi, e quale ritiene che siano le origini principali dei dati utilizzati dai membri del dipartimento?

4. Il/la candidato/a legga e traduca il seguente brano:

Multivariate analysis consists of a collection of methods that can be used when several measurements are made on each individual or object in one or more samples. We will refer to the measurements as variables and to the individuals or objects as units (research units, sampling units, or experimental units) or observations. In practice, multivariate data sets are common, although they are not always analyzed as such. But the exclusive use of univariate procedures with such data is no longer excusable, given the availability of multivariate techniques and inexpensive computing power to carry them out.

Historically, the bulk of applications of multivariate techniques have been in the behavioral and biological sciences. However, interest in multivariate methods has now spread to numerous other fields of investigation.

Quesiti busta 3

1. Il/la candidato/a illustri le linee guida dei Fair Data Principles a cui si ispira la policy di ateneo per il trattamento dei dati della ricerca

2. Si supponga che un docente dell'Università di Milano debba gestire ed analizzare dei dati di tipo socio-sanitario, nell'ambito di un progetto, con le seguenti caratteristiche:

    a. I dati sono suddivisi in 1000 file diversi, e sono in formati diversi: il 50% sono datasets in formato .csv, il 20% sono in formato .pdf (scansioni di documenti testuali, riportanti prescrizioni o diagnosi mediche) e il 30% sono in formato immagine (esiti di esami diagnostici). In totale occupano circa 10TB di memoria, ma possono essere trasferiti ed analizzati a gruppi di non più di 10 GB.
    b. I dati provengono da database sanitari di alcuni ospedali e contengono dati sensibili, che il docente è autorizzato a trattare, ma non a trasferire a terze parti. I dati andranno distrutti al termine del progetto.
    c. Le analisi che il docente deve eseguire includono analisi statistiche sui file .csv, tecniche di text mining sui documenti testuali e analisi di immagini sulle immagini mediche. Uno dei deliverables del progetto prevede anche di produrre un database anonimizzato contenente i risultati delle analisi effettuate.

    Il docente vorrebbe ridurre i tempi computazionali per le sue analisi, e vorrebbe utilizzare il miglior strumento di calcolo per ogni tipo di analisi da effettuare, e strutturare opportunamente il database richiesto dal progetto. Il docente non ha fondi disponibili per acquistare hardware, mentre ha fondi per acquistare servizi di calcolo.
    Quale architettura di calcolo, o quale insieme di architetture gli consiglierebbe di utilizzare per i suoi scopi? Si scelga tra un PC personale, attrezzature per HPC disponibili in ateneo (piattaforma Indaco), attrezzature disponibili presso centri di calcolo esterni (ad esempio Cineca), o servizi di cloud computing esterni all'ateneo, motivando la risposta.

3. Quali ritiene che siano le competenze necessarie ed i compiti di un Data Manager all'interno del Dipartimento di Economia, Management e Metodi Quantitativi, e quale ritiene che siano le origini principali dei dati utilizzati dai membri del dipartimento?

4. Il/la candidato/a legga e traduca il seguente brano:
    We use the term group to represent either a population or a sample from the population. There are two major objectives in separation of groups:
    1. Description of group separation, in which linear functions of the variables (discriminant functions) are used to describe or elucidate the differences between two or more groups. The goals of descriptive discriminant analysis include identifying the relative contribution of the p variables to separation of the groups and finding the optimal plane on which the points can be projected to best illustrate the configuration of the groups.
    2. Prediction or allocation of observations to groups, in which linear or quadratic functions of the variables (classification functions) are employed to assign an individual sampling unit to one of the groups. The measured values in the observation vector for an individual or object are evaluated by the classification functions to find the group to which the individual most likely belongs.
    For consistency we will use the term discriminant analysis only in connection with objective 1. We will refer to all aspects of objective 2 as classification analysis.

UNIVERSITÀ DEGLI STUDI DI MILANO

Milano, 20 luglio 2021

La Commissione

Prof.ssa Silvia Salini – Presidente

Prof.ssa Alessandra Micheletti – Componente

Dott.ssa Daniela Bagnati – Componente

Sig.ra Luisa Castellano - Segretaria