ROBERTO MARMO

Algoritmi per l'intelligenza artificiale

PROGETTAZIONE DELL'ALGORITMO
DATI E MACHINE LEARNING
NEURAL NETWORK • DEEP LEARNING

Web Pro

HOEPLI

ROBERTO MARMO

Algoritmi per l'intelligenza artificiale

Progettazione dell'algoritmo
Dati e Machine Learning
Neural Network – Deep Learning

EDITORE ULRICO HOEPLI MILANO

via Hoepli 5, 20121 Milano (Italy)
tel. +39 02 864871 – fax +39 02 8052886
e-mail hoepli@hoepli.it

www.hoeplieditore.it

ISBN 978-88-203-9171-3

Ristampa:

4 3 2 1 0 2020 2021 2022 2023 2024

Progetto editoriale: Maurizio Vedovati – Servizi editoriali (info@iltrio.it)

Stampa: L.E.G.O. S.p.A., stabilimento di Lavis (TN)

Printed in Italy

APPENDICE A

CONCETTO DI ALGORITMO

Definizione di algoritmo

Prima di cominciare a sviluppare una soluzione, è bene riflettere sul concetto di algoritmo, perché questo è il prodotto finale da realizzare e per evitare incomprensioni riguardo a un termine qualche volta abusato.

Un **algoritmo** è una sequenza finita e ordinata di operazioni non ambigue ed effettivamente computabili che, quando viene eseguita per risolvere un problema specifico, riceve un dato in ingresso (input), esegue delle elaborazioni, si arresta in un tempo finito e produce un dato come risultato (output). In fin dei conti, un algoritmo è una ricetta creata per risolvere un problema: si tratta di istruzioni semplici, ma basta seguirle alla lettera per fare cose ingegnose e complesse.

L'**euristica** è un concetto diverso da quello di algoritmo. Il termine deriva dal greco *heurískein*, "trovare, scoprire". Infatti, si definisce procedimento euristico un metodo di approccio alla soluzione dei problemi che non segue un chiaro percorso, ma si affida all'intuito e allo stato temporaneo delle circostanze, al fine di generare nuova conoscenza. In questo modo, può risolvere un problema più velocemente, nel caso in cui i metodi classici siano troppo lenti nel calcolo, o trovare una soluzione approssimata, nel caso in cui i metodi classici falliscano nel trovare una soluzione esatta. Il risultato viene ottenuto cercando di equilibrare gli obiettivi di un'ottimizzazione, una completezza, un'accuratezza e una velocità di esecuzione maggiori. L'inconveniente delle tecniche euristiche è dato dal fatto che esse si basano su decisioni che portano a scartare quella che può sembrare una strada poco promettente per il compito che si deve svolgere. Queste decisioni sono, in parte, intuitive, perciò, nel bene o nel male, condizionano notevolmente la qualità del sistema cui vengono applicate.

Proprietà dell'algoritmo

Le proprietà fondamentali di un algoritmo sono:

- non ambiguità: le istruzioni devono essere univocamente interpretabili dall'esecutore dell'algoritmo;
- eseguibilità: l'esecutore deve essere in grado, con le risorse a disposizione, di eseguire ogni istruzione in un tempo finito;

- finitezza: l'esecuzione di un algoritmo deve terminare in un tempo finito per ogni insieme di valori in ingresso; l'algoritmo deve quindi essere composto da un insieme finito di istruzioni (dopo ogni passo si sa precisamente qual è il successivo);
- ripetibile e deterministico, ossia produce lo stesso risultato a partire dagli stessi dati tutte le volte che viene eseguito;
- deve essere generale per produrre un risultato per ogni valore attribuito ai dati iniziali, cioè risolvere tutti i problemi della stessa classe;
- ciascuna istruzione deve essere eseguibile da parte dell'esecutore dell'algoritmo.

Giusto per dare un piccolo esempio, dato il problema di calcolare la soluzione di ax + b = 0 tramite la formula x = -b/a, si considera il seguente algoritmo:

- leggere i valori delle variabili a e b
- calcolare il valore -b
- dividere -b per a e assegnare il risultato alla variabile x
- scrivere sul monitor il contenuto della variabile x

Tutte le operazioni utilizzate per realizzare algoritmi rientrano in una delle seguenti tre categorie:

- operazioni sequenziali: un'istruzione sequenziale esegue una singola attività ben definita; terminata l'attività, l'algoritmo passa all'operazione successiva;
- operazioni condizionali: sono istruzioni in cui l'operazione successiva è selezionata sulla base della risposta fornita alla domanda contenuta;
- operazioni iterative: sono istruzioni che indicano di non proseguire con l'istruzione successiva, ma di tornare indietro e ripetere l'esecuzione di un precedente blocco di istruzioni.

Due importanti criteri per paragonare più algoritmi per lo stesso problema riguardano:

- efficienza temporale, cioè la capacità di avere risultati in un tempo accettabile;
- efficienza spaziale, cioè lo spazio necessario a memorizzare i dati deve essere accettabile.

Coding

Il linguaggio di descrizione delle operazioni deve essere adeguato alle caratteristiche del suo esecutore per poterlo realizzare concretamente. Un programma è un metodo computazionale (di calcolo) espresso secondo certe regole sintattiche cui sono associate precise azioni di un preciso esecutore; programmare significa soprattutto dividere la soluzione in azioni elementari fino al punto che ogni azione non si può suddividere in azioni più piccole e il computer può eseguire rapidamente l'azione. Quindi, un programma è la rappresentazione di un algoritmo con uno specifico linguaggio di programmazione, per quanto l'algoritmo stesso sia un costrutto intellettuale che esiste indipendentemente da qualsiasi rappresentazione.
Coding significa scrivere codice, ovvero programmare per tradurre i passi dell'algoritmo nei linguaggi di programmazione.

APPENDICE B
IMPARARE PYTHON

Questa appendice è utile per chi ha bisogno di imparare il linguaggio di programmazione Python e deve organizzare l'ambiente di programmazione per realizzare gli algoritmi mostrati nel libro *Algoritmi per l'intelligenza artificiale*.

Che cos'è Python

Si tratta di un linguaggio di programmazione interpretato ad alto livello. Sviluppato in ANSI C, Python è stato pubblicato da Guido van Rossum. Deriva il suo nome dalla commedia *Monty Python's Flying Circus* dei celebri Monty Python, in onda sulla BBC nel corso degli anni Settanta. Attualmente, lo sviluppo di Python viene gestito dall'organizzazione no-profit Python Software Foundation [1], grazie all'enorme e dinamica comunità internazionale di sviluppatori.
Python supporta diversi paradigmi di programmazione object-oriented con supporto all'ereditarietà multipla, imperativo e funzionale. Offre una tipizzazione dinamica forte. Fornito di una libreria built-in estremamente ricca, di gestione automatica della memoria e costrutti per la gestione delle eccezioni, è comodo, ha una sintassi pulita e snella e costrutti chiari; è semplice da usare e imparare perché è nato per essere un linguaggio immediatamente intuibile. Ha la caratteristica per cui i blocchi logici vengono costruiti semplicemente allineando le righe allo stesso modo, incrementando la leggibilità e l'uniformità del codice anche se vi lavorano diversi autori.
Python è un linguaggio pseudocompilato: un interprete si occupa di analizzare il codice sorgente contenuto in file testuali con estensione .py e, se sintatticamente corretto, di eseguirlo. In Python non esiste una fase di compilazione separata (come avviene in C, per esempio) che generi un file eseguibile partendo dal sorgente.
Per provare subito Python 3 si può andare su Trinket [2]; un'altra risorsa utile è Python Tutor [3], che permette di eseguire un'istruzione alla volta offrendo una utile visualizzazione di ciò che accade "sotto il cofano".

Python 2 e Python 3

Esistono alcune differenze tra le versioni che bisogna considerare, prima di scegliere quale usare o di decidere se passare a un'altra versione.
Pubblicato alla fine del 2000, Python 2 ha espresso un processo di sviluppo del linguaggio più trasparente e completo. Python 3 è considerato il futuro di Python, rilasciato alla fine del

2008 per affrontare e modificare i difetti di progettazione intrinseci delle precedenti versioni del linguaggio. L'obiettivo dello sviluppo di Python 3 è quello di ripulire il codice base e rimuovere la ridondanza, mettendo in chiaro che c'è un solo modo per eseguire un determinato compito. Python 3 è stato adottato lentamente, perché il linguaggio non è retrocompatibile con Python 2, obbligando gli utenti a prendere una decisione su quale versione del linguaggio utilizzare.
Dopo il rilascio di Python 3.0, nel 2010 è stato pubblicato Python 2.7, inteso come l'ultimo dei rilasci 2.x per rendere più facile agli utenti Python 2.x il porting di certe caratteristiche verso Python 3, fornendo un certo grado di compatibilità tra i due.
Su [4] vengono descritte le differenze sostanziali tra le due versioni; su internet si possono anche cercare le parole "differenze python 2 python 3". Le principali modifiche di Python 3.0 comprendono il cambiamento dell'istruzione print in una funzione built-in, il miglioramento del modo in cui gli interi sono divisi e un maggiore supporto a Unicode.

Versione usata nel libro *Algoritmi per l'intelligenza artificiale*

Salvo diversa indicazione, nel libro viene usato Python 3.0. Per passare il codice presentato alla versione Python 2.7 è necessario cambiare il formato dell'istruzione print, passando da, per esempio, print ("Ciao Mondo!") a print "Ciao Mondo!" tramite l'aggiunta delle parentesi tonde, perché in Python 3.0 l'istruzione print viene esplicitamente trattata come una funzione.
Nel sito web https://www.algoritmiia.it/ associato al libro è possibile trovare il codice nelle due versioni, se non ci sono problemi di incompatibilità con le librerie usate.

Perché usare Python

Esistono molti vantaggi derivanti dall'adozione di Python:

- è completamente gratuito ed è possibile usarlo e distribuirlo senza restrizioni di copyright;
- ampia comunità di sviluppatori per discussioni, aiuto, divulgazione;
- supporta la programmazione procedurale (che fa uso delle funzioni), a oggetti (includendo funzionalità come l'ereditarietà e altre tipicità), funzionale (come iteratori e generatori);
- lo stesso codice può essere eseguito su qualsiasi piattaforma avente l'interprete installato;
- è un linguaggio di alto livello ma semplice, intuitivo, facile da imparare; il comportamento del programma coincide con quanto ci si aspetta; si occupa automaticamente dell'allocazione e del rilascio della memoria;
- disponibilità di PyPy, un'implementazione altamente performante;
- ampia disponibilità di librerie e documentazione.

Un'altra caratteristica importante riguarda l'integrabilità con altri linguaggi. Infatti, oltre all'interprete classico scritto in C (denominato CPython per riferirsi al linguaggio usato per scrivere lo stesso interprete Python), esistono anche altri interpreti che consentono l'integrazione con diversi altri linguaggi. IronPython consente di utilizzare Python all'interno del framework .NET. Per integrare Python e Java è possibile utilizzare Jython.

Community

Cercando in https://www.meetup.com/it-IT/topics/python/it/ si può trovare un gruppo da frequentare vicino alla propria località.
PyCon [5] organizza un weekend per imparare, confrontarsi, approfondire la conoscenza.
Python.it [6] è il sito ufficiale della comunità italiana, con tanto materiale molto utile.

Corsi online

Sono centinaia le risorse disponibili online per imparare questo linguaggio. A titolo di esempio, le pagine di Html.it [7] o LearnPython [8]. Da considerare anche i vari MOOC illustrati nell'Appendice C.

Foglio di riepilogo

Un foglio di riepilogo è composto da una o due pagine, denominate anche cheat sheet, in cui vengono concentrati tutti i comandi più importanti da conoscere per una rapida consultazione senza dover cercare tra le pagine della documentazione. Per trovarli online si può cercare "cheat sheet python" per arrivare a pagine come [9] oppure [10].

Librerie da conoscere

NumPy è un'estensione che aggiunge supporto per vettori e matrici scaricabile da [11] per meglio gestire i calcoli matematici. Un utile foglio di riepilogo da tenere a portata di mano si trova su [12]. Per importare questa libreria si usa l'istruzione import numpy as np, dove np serve come alias per abbreviare numpy nella scrittura dei nomi delle funzioni da usare.
OpenCV [13] è una libreria software multipiattaforma nell'ambito della visione artificiale per elaborare immagini, riconoscere oggetti e tanto altro ancora, con documentazione su [14]. Per importare questa libreria bisogna scrivere il comando import cv2.
Pandas è la libreria più famosa per il data science descritta su [15] e nel libro di Molin; fornisce la base per importare e analizzare i dati, permette di ordinare le nostre infor-

mazioni in righe e in colonne, come se fossero array a più dimensioni di NumPy. Per importare questa libreria bisogna scrivere il comando import pandas as pd.
SciPy, disponibile su [16] con installazione spiegata in [17], è una collezione di algoritmi matematici e altre funzioni particolarmente utilizzate in ambito scientifico, a sua volta costruita su NumPy. Per importare questa libreria bisogna scrivere il comando import numpy as np per importare la libreria NunPy, seguito dal comando from scipy import nome per importare la libreria col nome indicato. Per conoscere questa libreria si può leggere il libro di Nunez-Iglesias.
Scikit-learn [18] è una libreria open source di apprendimento automatico per il linguaggio di programmazione Python. Contiene algoritmi di classificazione, regressione e clustering (raggruppamento) e macchine a vettori di supporto, regressione logistica, classificatore bayesiano, k-mean e DBSCAN, progettata per operare con le librerie NumPy e SciPy.

Da altri linguaggi a Python

Chi conosce il linguaggio R e vuole passare a Python può leggere il libro di Ajay Ohri. Cercando online "from R to Python", oppure "R to python conversion", si trovano pagine come [19] con spiegazioni su come passare all'altro linguaggio. Cercando "R and Python" oppure "Python and R" si trovano indicazioni come in [20] su come unire le forze di calcolo dei due linguaggi.
Con lo stesso meccanismo di ricerca, si possono trovare pagine web che spiegano come passare da Java e MatLab a Python.

Libri

La quantità di libri disponibili in lingua inglese è sterminata, in grado di coprire le basi e discorsi avanzati come hacking, alte prestazioni, interfacce grafiche, creazione di siti web, fino alla creazione di videogiochi.
Si può cominciare con un libro tascabile come quello di Beri. Un testo completo è fornito da Camagni, ricco di contenuti e apparati didattici che aiutano a comprendere, verificare con tanti esercizi, approfondire.

Anaconda

Anaconda è un gestore di pacchetti molto popolare perché semplifica sensibilmente il processo di setup di un ambiente di sviluppo per Python, racchiudendo assieme nella stessa distribuzione tutto ciò di cui si ha bisogno per iniziare subito a programmare. Bisogna cominciare con lo scaricare l'installer da [21].
In moltissimi libri dedicati a Python viene spiegato come attrezzare l'ambiente di programmazione. Si può cercare online "Installazione di Python utilizzando la distribuzio-

ne Anaconda" scrivendo di seguito il nome del sistema operativo. Una pagina utile si trova su [22]. Cercare i tutorial su YouTube è fondamentale per avere un aiuto visivo.

Jupiter

Jupyter è un notebook, ovvero un'applicazione web gratuita e open source che funziona su un'applicazione client basata sul web e si avvia con un browser standard. Permette di creare e condividere documenti che contengono codice in Python, equazioni, testi e immagini incorporate, ovvero rende possibile unificare strumenti finora normalmente separati del lavoro di uno sviluppatore. Si possono unire le fasi di raccolta dei dati, scrittura del codice, visualizzazione di grafici e tabelle nonché la possibilità di rendere il codice condivisibile sulla piattaforma GitHub.
Il codice è, di volta in volta, modificabile ed eseguibile in tempo reale. Quanto realizzato può essere esportato in formato HTML, PDF e LaTeX. Si possono scrivere formule matematiche anche complicate tramite il formato LaTeX.
Istruzioni sull'uso si possono trovare su [23] e [24].
Viene creato un file nel formato .ipynb in grado di memorizzare tutto il materiale scritto con il formato adatto all'interscambio JSON (JavaScript Object Notation).
Volendo cercare qualcosa su Google per provare subito a eseguire del codice, si possono scrivere le parole chiave e la specifica filetype:ipynb, per esempio: machine learning filetype:ipynb.

Google CoLab

Google Colab è una piattaforma online gratuita che offre un servizio di cloud hosting di Jupyter Notebooks su [25]. Richiede un account in Google, come quello per gestire la posta con Gmail. Supporta Python 2.7 e 3.6, librerie come PyTorch, Tensorflow, Keras e OpenCv, ed è possibile installare altri moduli. Potrebbe arrivare anche il supporto per R e Scala. L'ambiente permette di salvare tutto il codice creato su Google Drive ed esportare su GitHub.
Il punto di forza è la disponibilità anche del supporto gratuito per GPU e le nuove TPU. In tal modo, si comincia subito a programmare senza impegnarsi nell'installazione di software e costosi hardware nel proprio computer. Un articolo sull'attivazione della GPU si trova su [26].
Le FAQ si trovano su [27], un utile articolo su come avviare la programmazione su [28]; la pagina di benvenuto [29] spiega il servizio grazie anche a un breve video, altre pagine con spiegazioni sull'uso si trovano su [30], [31] e [32].
Occorre prestare attenzione al modo in cui si salvano i file nella sezione a sinistra, per evitare di non trovarli al prossimo ritorno e doverli ricaricare. Conviene inserirli nella cartella sample_data. Questa sezione si può aggiornare cliccando su Refresh per mostrare i nomi dei file caricati.

Per trovare online dei documenti Colab disponibili da provare subito si può provare una ricerca del tipo: parole chiave site:https://colab.research.google.com, per esempio: "demo machine learning site:https://colab.research.google.com"; conviene prestare attenzione alla fonte che li ha scritti per evitare di trovare materiale con errori.
Installare una libreria da un repository come https://github.com/ è molto facile. Basta andare sull'indirizzo di interesse, cliccare su "Clone or download" per ottenere l'indirizzo https, copiare l'indirizzo, per esempio: https://github.com/DEAP/deap.git. Nell'interfaccia di Colab si scrive il comando `!git clone https://github.com/DEAP/deap.git` in cui si può notare il ! prima del comando, seguito da `!pip install deap` per completare l'installazione della libreria, come indicato nella figura successiva.
Le FAQ su [33] rispondono alle domande più frequenti. Notare la sezione in fondo alla pagina in cui viene annunciato che dal 1° gennaio 2020 è stato tolto il supporto a Python 2.
Per trovare altri esempi su come installare le librerie si possono cercare su internet le parole: "import Machine learning libraries into Google Colab".
Nella Figura B.1 si possono notare i vari elementi dell'interfaccia grafica. A sinistra, i file caricati e in alto i comandi per aggiungere le celle con testo o codice; nelle singole celle si nota la freccia bianca nel cerchio grigio per avviare l'esecuzione del codice contenuto.

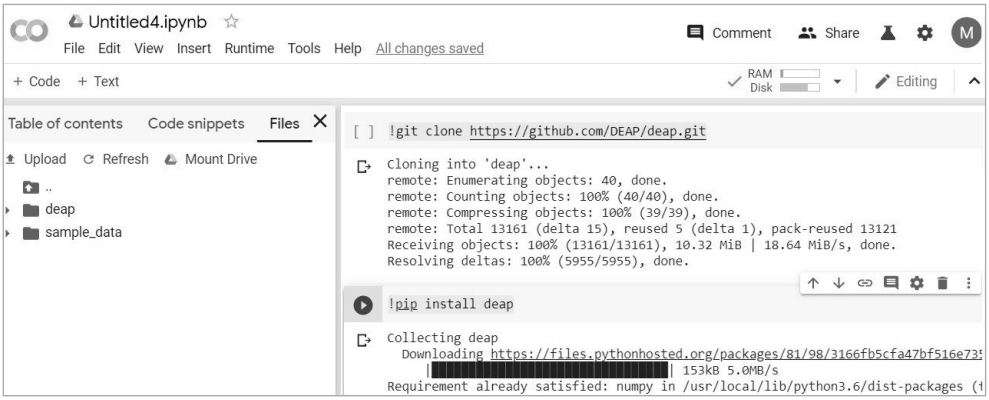


Figura B.1 – Interfaccia di Google Colab per programmare subito in linguaggio Python, esempio di installazione di una libreria.

Per caricare in Colab un file esistente in Google Drive bisogna usare le istruzioni spiegate in [34] e [35]; conviene collegarli per evitare di perdere i file caricati in Colab. La sessione dura al massimo 12 ore, poi variabili e file vengono cancellati e bisogna ricominciare.

Riferimenti bibliografici

M. Beri, *Guida tascabile al linguaggio di Google, Star Wars e la NASA*, Apogeo, Milano 2010.
P. Camagni, R. Nikolassy, *Python – Quaderni di tecnologie*, Hoepli, Milano 2019.
S. Molin, *Hands-On Data Analysis with Pandas: Efficiently perform data collection, wrangling, analysis, and visualization using Python*, Packt Publishing, Birmingham (UK) 2019.
J. Nunez-Iglesias, *Elegant SciPy: the art of scientific Python*, O' Reilly, Sebastopol (CA) 2017.
A. Ohri, *Python for R Users: A Data Science Approach*, John Wiley & Sons, Hoboken (NJ) 2018.

Note

1 Python Software Foundation, http://www.python.org/psf/
2 Python 3 Trinkets, https://trinket.io/features/python3
3 Python Tutor, http://pythontutor.com/visualize.html#mode=edit
4 *Python 2 vs Python 3: considerazioni pratiche*, https://www.0x90.it/python-2-vs-python-3/
5 Pycon, https://www.pycon.it/it/
6 Python.it, http://www.python.it/
7 Guida Python di Html.it, https://www.html.it/guide/guida-python/
8 Learn Python, https://www.learnpython.org/
9 Python Cheatsheet, https://www.pythonforbeginners.com/cheatsheet/ https://www.pythoncheatsheet.org/
10 List of Data Science Cheatsheets to rule the world, https://github.com/FavioVazquez/ds-cheatsheets#python
11 Libreria NumPy, http://numpy.scipy.org/
12 Python For Data Science Cheat Sheet NumPy Basics, https://s3.amazonaws.com/assets.datacamp.com/blog_assets/Numpy_Python_Cheat_Sheet.pdf
13 Libreria OpenCV, https://opencv.org/
14 Documentazione OpenCV, https://www.learnopencv.com/ con Python https://opencv-python-tutroals.readthedocs.io/en/latest/
15 Python Data Analysis Library, https://pandas.pydata.org/
16 Libreria Scipy, https://www.scipy.org/ https://scipy-lectures.org/
17 Scipy installation, https://www.scipy.org/install.html
18 Scikit-learn, https://scikit-learn.org/
19 *R'un towards Python,* https://www.kaggle.com/hiteshp/r-to-python-tutorial
20 *From "R vs Python" to "R and Python"*, https://towardsdatascience.com/from-r-vs-python-to-r-and-python-aa25db33ce17
21 https://www.anaconda.com/download
22 https://softpython.readthedocs.io/it/latest/installation.html
23 *Notebook Jupyter: documenti web per analisi di dati, livecode e molto altro*, https://www.ionos.it/digitalguide/siti-web/programmazione-del-sito-web/notebook-jupyter/
24 *Jupyter Notebook for Beginners: A Tutorial*, https://www.dataquest.io/blog/jupyter-notebook-tutorial/
25 Google CoLab, https://colab.research.google.com
26 *Google Colab Free GPU Tutorial*, https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d
27 Colaboratory Frequently Asked Questions, https://research.google.com/colaboratory/faq.html

28 *Come addestrare una Intelligenza Artificiale... Gratis! Trucchi e consigli pratici*, https://medium.com/@cosimo.iaia/come-addestrare-una-intelligenza-artificiale-gratis-trucchi-e-consigli-pratici-f37935916f3a
29 Welcome to Colaboratory!, https://colab.research.google.com/notebooks/welcome.ipynb#
30 Google Colaboratory - Reti Neurali Artificiali, https://www.retineuraliartificiali.net/google-colaboratory/
31 *Getting Started With Google Colab. A Simple Tutorial for the Frustrated and Confused*, https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c
32 *How to use Google Colab*, https://www.geeksforgeeks.org/how-to-use-google-colab/
33 *Google Colab FAQ*, https://research.google.com/colaboratory/faq.html
34 External data: Local Files, Drive, Sheets, and Cloud Storage, https://colab.research.google.com/notebooks/io.ipynb
35 *Downloading Datasets into Google Drive via Google Colab*, https://towardsdatascience.com/downloading-datasets-into-google-drive-via-google-colab-bcb1b30b0166

APPENDICE C
RISORSE PER STUDIARE

Ogni volta che impariamo qualcosa di nuovo, noi stessi diventiamo qualcosa di nuovo.
Leo Buscaglia

Questa appendice fornisce i riferimenti online alle migliori risorse per studiare AI, trovare algoritmi, trovare altre persone interessate. Vengono discussi i principali riferimenti; sul sito web https://www.algoritmiia.it/ associato al libro *Algoritmi per l'intelligenza artificiale* sono disponibili ulteriori link.

Associazioni

Iscriversi a un'associazione di settore porta molti vantaggi in cambio di qualche decina di euro: aggiornamenti, partecipazione a congressi, possibilità di conoscere altre persone, newsletter, un certo senso di appartenenza a un gruppo ecc.

AIxIA

L'Associazione Italiana per l'Intelligenza Artificiale (AIxIA) [1] è un'associazione scientifica senza fini di lucro, fondata nel 1988 con lo scopo di promuovere la ricerca e la diffusione delle tecniche proprie dell'intelligenza artificiale. Promuove lo studio e la ricerca sull'AI e coordina le attività del settore in Italia. Organizza conferenze scientifiche, eventi di disseminazione e approfondimento e forum aziendali su temi relativi. Mantiene mappe dell'ecosistema italiano, della ricerca italiana e dei suoi membri nonché dei corsi universitari.
L'Associazione si pone l'obiettivo di aumentare la conoscenza dell'intelligenza artificiale, incoraggiarne l'insegnamento e promuovere la ricerca teorica e applicata nel campo attraverso seminari, iniziative mirate e sponsorizzazione di eventi.
In seno all'Associazione sono presenti gruppi di lavoro focalizzati su temi di ricerca specifici. Al momento sono attivi otto gruppi di lavoro: sistemi ad agente e multi-agente, intelligenza artificiale e ageing, intelligenza artificiale per i beni culturali, robotica, apprendimento automatico e data mining, rappresentazione della conoscenza e ragionamento automatico, elaborazione del linguaggio naturale, argomentazione.
L'Associazione, che attualmente conta oltre 1000 membri, organizza un evento scientifico annuale e iniziative dirette al pubblico e al mondo industriale; offre premi e borse di studio per favorire la partecipazione degli studenti e dei giovani ricercatori agli eventi che si tengono in Italia.

L'AIxIA è membro della European Association for Artificial Intelligence EurAI (precedentemente ECCAI). EurAI è stata fondata nel 1982 per rappresentare la Comunità Europea dell'Intelligenza Artificiale.

CVPL

Il CVPL (ex GIRPR) [2], Associazione italiana per la ricerca in Computer Vision, Pattern Recognition e Machine Learning, è nata nel 1983 ed è affiliata all'International Association in Pattern Recognition (IAPR).
La principale mission del CVPL è creare e mantenere attiva la comunità scientifica italiana nelle discipline della Computer Vision e Image Processing, Pattern Recognition e Machine Learning, che in questi anni sono tra i temi più caldi della ricerca internazionale nell'informatica e nell'ingegneria informatica e di grande interesse per l'industria ICT.
Il CVPL si occupa di argomenti teorici, dai modelli a grafi al deep learning, su dati immagini, video e 3D, sul colore, le tessiture e il movimento, e su dati sensoriali differenti, ora anche acquisiti in IoT. La ricerca poi si declina in molti temi applicativi correlati, quali document analysis, medical imaging, videosorveglianza, biometria, multimedia, cultural heritage, automotive, visione robotica, interazione uomo-macchina e applicazione verso la grafica e l'augmented reality.
Il CVPL supporta e favorisce l'internazionalizzazione della ricerca italiana, particolarmente nei confronti dello IAPR (più di 20 membri CVPL sono Fellow IAPR), ed è anche collegato ad altre associazioni quali ACM SIGMM, IEEE Biometric Society, Computer Vision Foundation.
Il CVPL promuove iniziative di diffusione scientifica: la conferenza ICIAP biennale, il convegno CVPL e conferenze e workshop locali su tutto il territorio italiano; sponsorizza inoltre iniziative di formazione, scuole estive e tutorial come VISMAC, scuola biennale attiva da oltre 20 anni, e gestisce premi per la ricerca scientifica soprattutto nei confronti dei dottorandi e dei giovani ricercatori.
Il CVPL ha formato generazioni di ricercatori in pattern recognition e discipline affini che hanno un'attiva carriera accademica, nei centri di ricerca e nell'industria. È fonte di aggregazione per progetti nazionali e internazionali di ricerca teorica e applicata.
Il CVPL conta più di 300 soci attivi e altre centinaia di ricercatori di accademia e industria che seguono le attività nelle diverse iniziative.

IAML

La Italian Association for Machine Learning (IAML) [3] ha come scopo principale la promozione e la diffusione di studi, iniziative, ricerche, informazioni e aggiornamenti in materia di machine learning, intelligenza artificiale e argomenti affini; l'organizzazione di seminari, eventi, corsi di formazione, qualificazione e aggiornamento professionale, convegni e incontri di approfondimento sulle medesime tematiche; la realizza-

zione e la diffusione di contributi editoriali, in forma elettronica e cartacea; la sensibilizzazione dei consumatori su tematiche di interesse giuridico, economico e tecnologico; la promozione e la gestione di attività editoriali in genere, informative e di comunicazione e delle relative attività accessorie.

SIREN

La Società Italiana Reti Neuroniche (SIREN) [4], fondata nel 1989, si occupa di Reti Neurali Artificiali. Creata da un'intuizione del professor Eduardo Caianiello, che nel 1988 radunò a Vietri alcuni tra i più rappresentativi scienziati italiani nel campo delle neuroscienze, trovò ospitalità presso la sede dell'IASS a Vietri sul Mare, dove ogni anno organizza il convegno internazionale WIRN che mette a confronto i gruppi di ricerca italiani con i più significativi ricercatori internazionali.

Blog

Esistono tantissimi blog in lingua inglese, su temi generali e specifici. Una prima scelta porta verso blog gestiti da una persona [5] oppure da strutture più grandi come Data Science Dojo [6], IBM [7], Towards Data Science [8], Kdnuggets [9], Data Science Central [10].
Per trovare i migliori blog si possono usare queste combinazioni di parole: Top Artificial Intelligence Websites And Blogs for AI Enthusiast [11], Top Active Blogs on AI [12], Blogs and bloggers to follow [13], Must-read AI Machine Learning blogs [14], The Best AI and Machine Learning Blogs to Follow [15].

Centri di ricerca

I centri di ricerca presso le università svolgono attività di sviluppo della ricerca scientifica e creazione di soluzioni. Organizzano anche attività di divulgazione della conoscenza tramite eventi e pubblicazione di materiali sui loro siti internet; ogni tanto conviene visitare le loro pagine.

Consorzio CINI

Il CINI (Consorzio Interuniversitario Nazionale per l'Informatica) [16] costituisce oggi il principale punto di riferimento della ricerca accademica nazionale nei settori dell'informatica e dell'Information Technology. Costituito il 6 dicembre 1989, il CINI è posto sotto la vigilanza del Ministero competente per l'università e la ricerca, include solo università pubbliche e costituisce soggetto in house rispetto agli enti costitutori, partecipanti e legittimamente affidanti. Non ha scopo di lucro, né può distribuire utili.
Il Consorzio si è sottoposto alla Valutazione della Qualità della Ricerca da parte dell'ANVUR. È costituito da 47 università pubbliche ed è attualmente dotato di 10 la-

boratori nazionali, con oltre 1300 docenti afferenti ai settori scientifico-disciplinari INF/01 e ING-INF/05.
Il Consorzio promuove e coordina attività scientifiche, di ricerca e di trasferimento, sia di base sia applicative, nel campo dell'informatica, di concerto con le comunità scientifiche nazionali di riferimento. Favorisce, in particolare:

- la collaborazione con università, istituti di istruzione universitaria, enti di ricerca, aziende e pubblica amministrazione;
- l'accesso e la partecipazione a progetti e attività scientifiche, di ricerca e di trasferimento;
- la creazione e lo sviluppo di laboratori tematici nazionali;
- la realizzazione di percorsi di alta formazione.

In tutte le attività, il CINI è in grado di garantire:

- la massima qualità a livello nazionale (e, ove necessario, internazionale), potendo attingere alle varie eccellenze accademiche;
- la massa critica necessaria al raggiungimento degli obiettivi concordati;
- la distribuzione geografica su tutto il territorio nazionale.

A livello internazionale, il CINI è membro dell'Executive board della BDVA (Big Data Value Association), che ha lanciato la parte privata nella cPPP (Contractual Public Private Partnership) sui Big Data Value; partecipa alle attività di Ecsel JU, Artemis JTI, NESSI (Networked European Software and Services Initiative); tramite il National Expert Group, supporta attivamente il delegato italiano per il Comitato ICT del programma Horizon 2020. È membro attivo dell'European Forum for ICST (EFICST) e di Informatics Europe (IE).
A livello nazionale, il CINI, grazie ad accordi quadro, è coinvolto in progetti di ricerca, trasferimento tecnologico e di alta formazione con i principali player del sistema industriale nazionale e con consorzi sia pubblici sia privati; collabora con le principali associazioni nazionali dei professionisti dell'ICT.

IIT

L'Istituto Italiano di Tecnologia (IIT) [17] è stato fondato a Genova nel 2003; è una fondazione disciplinata dal Codice civile e finanziata dallo Stato per lo svolgimento di attività di ricerca scientifica di interesse generale, per fini di sviluppo tecnologico. È vigilato dal Ministero dell'economia e delle finanze e dal Ministero dell'istruzione, università e ricerca e sottoposto al controllo della Corte dei conti. La Fondazione intende promuovere lo sviluppo tecnologico e la formazione avanzata nel paese, in accordo con le politiche nazionali a favore della scienza e della tecnologia, rafforzando così il sistema di produzione nazionale. Tra i settori di punta delle ricerche sul fronte dell'intelligenza artificiale vi è anche quello dedicato all'analisi di modelli di visione artificiale, robotica, neuroscienze e tecnologie del cervello. Qui è nato iCub (I come in "I Robot" e Cub come "cucciolo d'uomo", *man-cub*, dal *Libro della giungla* di Kipling), il robot uma-

noide con funzioni cognitive e dimensioni corporee simili a quelle di un bambino di cinque anni, progettato specificamente per supportare la ricerca nel campo dell'intelligenza artificiale, e in grado di gattonare, camminare e sedersi per manipolare oggetti.

Laboratorio CVML a Pavia

Il laboratorio di Computer Vision and Multimedia Lab (CVML) [18] del Dipartimento di Ingegneria industriale e dell'informazione presso l'Università di Pavia è impegnato in attività di ricerca concentrate sull'elaborazione delle immagini, architetture parallele per la visione artificiale, interfacce grafiche usabili, studio del social media marketing. Attualmente è attivo nelle seguenti aree di ricerca: Riconoscimento di persone, Deep Reinforcement Learning per la robotica, Proteomica, Eye Tracking per l'interazione uomo-macchina e la biometria, Digital Humanities con l'impiego del 3D per promuovere la divulgazione e l'accessibilità del patrimonio culturale. In quest'ultimo ambito, in particolare, sono state realizzate riproduzioni 3D e didascalie tattili di importanti manufatti artistici, per percorsi sensoriali fruibili durante esposizioni temporanee e permanenti e integrabili in progetti per lo sviluppo di applicazioni di realtà aumentata.

Codice di programmazione

Trovare listati di codice su internet è molto facile, grazie alla grande varietà delle fonti, ma bisogna stare attenti a varie problematiche: la tutela della proprietà intellettuale, la correttezza nella scrittura, la chiarezza del funzionamento.

GitHub

Il controllo versione (version control), in informatica, è la gestione di versioni multiple di un insieme di informazioni. Gli strumenti software per il controllo versione sono ritenuti molto spesso necessari per la maggior parte dei progetti di sviluppo software.
Git [19] è lo strumento più popolare per lavorare con i sistemi di controllo di versione del software. Creato da Linus Torvalds, conosciuto soprattutto per essere stato l'autore della prima versione del kernel Linux e coordinatore del progetto di sviluppo dello stesso, Git è un software open source ed è un ottimo strumento di cooperazione grazie alla sua natura distribuita che consente operazioni veloci e permette una gestione avanzata. Attorno a esso è stato costruito GitHub [20], che ne ha valorizzato le potenzialità rendendole accessibili via web. Infatti, GitHub è un servizio di hosting per progetti software con distribuzione gratuita; il nome deriva dal fatto che è una implementazione dello strumento di controllo versione distribuito Git. Sistemi come questi vengono chiamati anche "repository di codice".
Per imparare a usare GitHub si può leggere il libro di Pipinellis e pagine di siti web specializzati [21] [22], in particolare il tutorial [23] creato all'interno di GitHub, un esempio di utilizzo per la pubblicazione di pagine web.

L'operazione di download viene detta clone; serve identificare il percorso del repository che volete scaricare. Per esempio, considerando la versione associata al libro *Algoritmi per l'intelligenza artificiale* su https://www.algoritmiia.it/ illustrata in Figura C.1, in basso a destra si trova il pulsante verde "Clone or download"; cliccare per far comparire una piccola finestra bianca in cui cliccare su Download ZIP per ottenere una cartella compressa formato ZIP contenente tutto il materiale fornito dallo specifico repository. Per trovare qualcosa di interessante, si può partire dalla ricerca in Google con eventuale aggiunta della funzione site:github.com per cercare soltanto in questi repository.

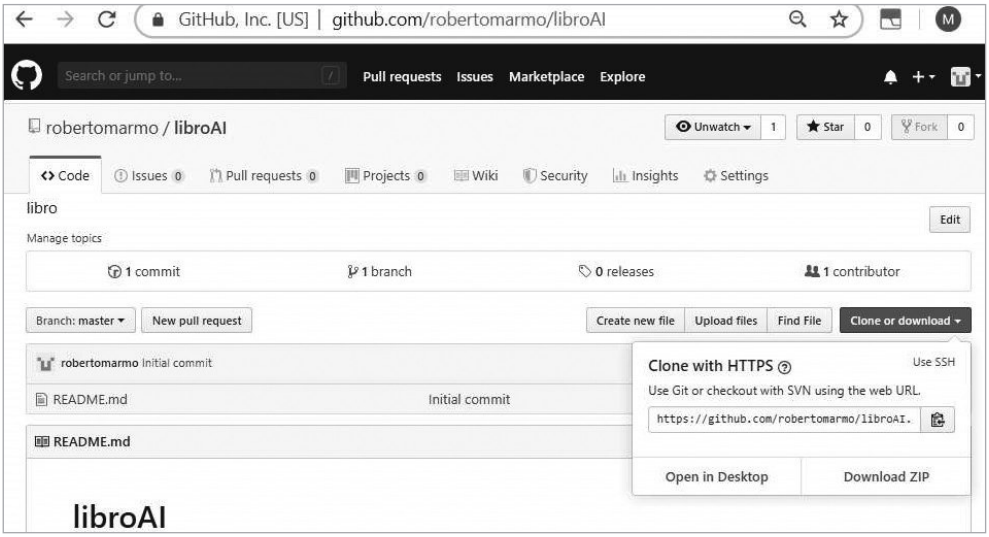


Figura C.1 – Scaricare codice da GitHub con il pulsante verde in basso a destra.

Ricerca con Google

L'opzione filetype indica a Google quale tipo di file deve cercare; si dimostra molto utile per trovare documenti, per esempio filetype:pdf per documenti in formato Adobe PDF o filetype:docx per formato Microsoft Word, o file con il codice, per esempio: filetype:py per listati in Python, filetype:ipynb per IPython Notebook o Jupyter Notebook.

L'opzione site indica in quale sito specifico bisogna cercare, per restringere la ricerca ed evitare troppi risultati inutili; per esempio, si può cercare nel servizio Colab di Google (spiegato nell'Appendice B) per trovare molti tutorial sul machine learning: machine learning site:colab.research.google.com.

Ovviamente, vanno aggiunte le varie parole necessarie per individuare cosa bisogna trovare e si possono usare varie funzioni contemporaneamente, per esempio: deep learning tutorial filetype:ipynb site:colab.research.google.com.

Siti web dei libri

Quasi tutti i libri di programmazione permettono di scaricare il codice discusso nelle loro pagine, attraverso il sito web dell'editore, il sito web dell'autore, sistemi come GitHub e simili. I riferimenti ai link si trovano nelle pagine di introduzione. Con un motore di ricerca online si possono aggiungere parole come "book code" oppure "download the code file" insieme alle parole che descrivono il codice desiderato. Conviene aggiungere anche qualche nome di editore specializzato.
Per trovare i link all'interno del libro in formato digitale conviene cercare la parola http e scorrere i riferimenti trovati; provare anche con "download the code file".

Community e forum

Si tratta della versione digitale delle piazze in cui si incontrano le persone per discutere intorno ad argomenti di interesse. Nelle community è più facile conoscere nuove persone, grazie agli strumenti di socializzazione, mentre i forum sono costruiti soprattutto sulla discussione. Per i riferimenti ai social media si veda uno dei prossimi paragrafi.

AI Italia

Community libera di esperti, consulenti, studiosi e appassionati di tutto ciò che rientra nell'ambito dell'AI, realizzata e gestita da Fluel.it [24], che raggruppa le attività di consulenza sul tema Artificial Intelligence e Disruptive Innovation, nonché gli eventi formativi e di networking, dedicati a C-Level, Manager e Decision Maker sulle più importanti tecnologie e normative emergenti. Sono disponibili il sito web [25], Telegram [26], gruppi di discussione in Facebook [27] e LinkedIn [28].

Data Science Milan

Community di data scientist ed esperti di machine learning localizzata nell'area di Milano [29]. È un gruppo indipendente con l'unico obiettivo di promuovere la conoscenza e l'innovazione dell'AI e della rivoluzione basata sui dati in Italia, e non solo. Consente di ottenere facilmente risposte a domande specifiche, oltre a fornire accesso a ricercatori, influencer, esperti e guide che possono aiutare a orientarsi in una galassia di metodi e tecnologie. Presente su Facebook [30] e LinkedIn [31] e [32].

Datascienceseed

Datascienceseed [33] è un gruppo indipendente e no-profit di appassionati di machine learning e data science; parte della già citata IAML, coltiva queste competenze con meetup e gruppi di studio a Genova.

Deep Learning Italia

La più grande community dedicata al deep learning, cui offre articoli, meetup, tutorial, lezioni, consulenze. Nella homepage si presenta così: "Tutte le informazioni necessarie per diventare un esperto di Deep Learning al fine di analizzare i Big Data nella tua attività". Su Facebook sono disponibili il gruppo e la pagina [34], LinkedIn [35], Meetup [36].

Kaggle

Kaggle [37] è una comunità online di scienziati e studenti informatici, di proprietà di Google LLC. Kaggle consente agli utenti di trovare e pubblicare articoli e set di dati, di esplorare e costruire modelli. Ospita anche competizioni di machine learning [38].

ODSC

ODSC [39] è una community dedicata al data science sviluppata in tutto il mondo, per impegnarsi, costruire, sviluppare e imparare. Una sezione composta da volontari organizza eventi a Milano con un gruppo Meetup [40]. Da notare la sezione dedicata alla formazione [41] in cui rivedere importanti webinar.

Stack Overflow

Stack Overflow [42] è un sito web che fa parte della rete Stack Exchange in cui si possono inserire domande e ricevere risposte riguardo a vasti argomenti di programmazione. È disponibile solo in lingua inglese. Spesso viene pubblicato del codice di cui si chiede la correzione, oppure si ottiene il codice in seguito a una domanda su come fare qualcosa.

La community riscuote in pochissimo tempo talmente tanto successo da instillare negli ideatori, Jeff Atwood e Joel Spolsky, l'idea di creare tante community fondate sul medesimo principio di Stack Overflow ma dedicate a tematiche differenti. Su [43] si trovano le statistiche sulle tipologie di utenti.

Ogni domanda e ogni risposta ottengono un punteggio da parte degli utenti, facendo sì che ciascun partecipante possa ottenere una valutazione. Contiene anche offerte di lavoro [44].

Stack Overflow può essere usato come fonte per capire cosa succede nel settore AI; per esempio, [45] fa notare come Python sia sempre tra i linguaggi più ricercati su Stack Overflow, una delle piazze più frequentate da persone interessate all'AI: da ciò si può dedurre l'importanza di Python nel fare AI.

Si può trovare talmente tanta informazione da non riuscire subito a individuare ciò di cui si ha bisogno; è quindi importante indicare molte parole per specificare cosa si sta cercando. Google raggiunge le sue pagine, quindi spesso le fornisce nel risultato di ricerca, oppure si può indirizzare la ricerca con site:stackoverflow.com.

The Data Literacy Project

Si tratta di una community [46] impegnata a permettere a chiunque di comprendere, analizzare e utilizzare i dati. Nel forum individui e organizzazioni possono interagire con leader e influencer del settore per condividere idee in grado di alimentare l'alfabetizzazione dei dati. Data Literacy consente a tutti di fare domande sui dati, prendere decisioni e comunicare le proprie idee agli altri.

Corsi universitari

I percorsi di laurea in statistica, matematica, ingegneria dell'informazione, informatica e similari offrono spesso un corso di intelligenza artificiale, facoltativo oppure obbligatorio. Nella guida per lo studente vengono riportati i corsi di riferimento; si può cercare anche nel sito web del corso di laurea raggiungibile dalla home page del sito dell'università preferita.

Eventi pubblici

Partecipare agli eventi pubblici ha importanti vantaggi: aumentare le conoscenze tecniche, incontrare nuove persone, farsi conoscere se si partecipa come relatori. Esistono eventi nel mondo reale, dove le persone si incontrano fisicamente, ed eventi online tramite internet, con minore possibilità di relazione diretta ma cui si può partecipare comodamente da casa.
Una semplice ricerca su internet con parole quali "evento argomento località" può permettere di trovare link interessanti sull'argomento trattato in un evento che si svolge in una certa località.

Meetup

Meetup è un servizio di rete sociale che ha lo scopo di facilitare l'incontro di gruppi di persone in varie località del mondo. Il funzionamento è descritto su [47], la pagina italiana è [48]. Meetup consente ai suoi membri di trovare e unirsi a gruppi creati attorno a un comune interesse, come la politica, i libri, i giochi ecc. Un esempio di evento si trova su [49]. Si può cercare per argomento o per località. Essendo una rete sociale, ogni partecipante può creare il proprio profilo personale, utile anche per aumentare la propria presenza su internet; ne è un esempio quello creato da Roberto Marmo, autore del libro *Algoritmi per l'intelligenza artificiale* [50]. In ogni pagina di evento si possono trovare i riferimenti ai profili dei partecipanti, con cui capire meglio chi partecipa. Meetup può essere un'occasione per creare nuove relazioni professionali; quando qualcuno indica l'azienda in cui lavora, questo può servire per capire chi sono i dipendenti di una certa azienda e in cosa sta investendo. Spesso in questa pagina si trovano link ai materiali presentati, così si possono trovare slide interessanti anche se non si è potuto

partecipare all'evento relativo. Il riepilogo dei gruppi italiani sulla data science si trova su [51] e [52] per AI, alcuni gruppi interessanti sono [53] [54] [55] [56] [57] [58].

Eventbrite

Eventbrite è un servizio online di gestione e biglietteria di eventi con sede negli Stati Uniti. Esso consente agli utenti di navigare, creare e promuovere eventi locali. Il servizio addebita una commissione agli organizzatori di eventi in cambio di servizi di biglietteria online, a meno che l'evento non sia gratuito. Si può usare il motore di ricerca interno per trovare qualcosa secondo l'area geografica o l'argomento di interesse. Il riepilogo degli eventi sul machine learning a Milano si trova su [59] e per AI su [60]. Entrando nel proprio account si trovano i biglietti per gli eventi e ulteriori suggerimenti.
Basta iscriversi a uno dei servizi precedenti per ricevere email con informazioni sugli eventi successivi, oppure si può leggere qualche pagina dei siti web per trovare altri suggerimenti.
Chi frequenta le università può ricevere la newsletter con gli eventi; può anche passare davanti alle bacheche per leggervi i foglietti appesi con gli annunci.
Certamente il numero di eventi intorno a un certo tema è un indice significativo della sua importanza; anche la località indica in quale area geografica esso è più o meno importante.

Foglio di riepilogo

Un foglio di riepilogo è un insieme conciso di note utilizzato per una rapida consultazione, quando non si ricorda qualcosa e non si ha tempo per cercare tra le tante pagine di libri e siti web. I fogli di riepilogo sono chiamati anche cheat sheet perché possono essere utilizzati dagli studenti per imbrogliare su un test.
Spesso vengono plastificati, perché non si rovinino durante l'uso frequente.
Conviene dargli un'occhiata quando vengono scaricati per la prima volta, in modo da individuare quali informazioni sono disponibili e come raggiungerle in poco tempo.
Spesso vengono rilasciati nel formato Adobe PDF, per cui si può facilmente cercare con la funzione Trova per andare subito a ciò che serve.
Su [61] e [62] una vasta raccolta di link verso fogli con il riepilogo dei principali algoritmi permette di comprenderne il meccanismo di funzionamento grazie all'ampio uso di colori e testi.

Infografiche

L'infografica (anche nota con i termini inglesi "information design", "information graphic" o "infographic") è l'informazione proiettata in forma più grafica e visuale che testuale. Come tecnica è nata dall'incrocio della grafica con il giornalismo. In un solo pannello occupante poco spazio si possono trovare dati statistici, diagrammi di flusso,

riepiloghi, liste e tanto altro ancora. Si tratta di un riassunto visuale, molto efficace per la comprensione e per la memoria visiva.
Pinterest.com è un social network basato sulla condivisione di fotografie, video e immagini. Il nome deriva dall'unione delle parole inglesi "pin" e "interest". Pinterest permette agli utenti di creare bacheche in cui catalogare le immagini presenti nelle pagine web in base a temi predefiniti oppure da loro scelti. Esistono tante bacheche interessanti, per esempio [63].

Libri

Cercando all'interno di siti di e-commerce come Amazon e di quelli di editori come Hoepli è possibile trovare titoli e relativi autori, recensioni dei lettori e altri libri suggeriti. Molto spesso i libri sono accompagnati da un sito web, dove trovare approfondimenti e il codice di programmazione presentato.
Gli studenti universitari possono usare il sito web della biblioteca del proprio ateneo e chiedere libri in prestito da altre biblioteche.
Per sapere in quale biblioteca si trova un libro si può usare il catalogo del servizio bibliotecario nazionale [64].
Google Libri o Google Ricerca Libri è l'interfaccia in italiano di Google Books [65], lo strumento sviluppato da Google per permettere la ricerca nel testo di libri antichi digitalizzati oppure in commercio. Secondo il copyright associato al libro, consente di visualizzare piccole porzioni del libro in cui si trova il testo cercato. Basta usare il motore Google e guardare nelle prime righe dei risultati, oppure l'indirizzo specifico.

Master

I master di primo e secondo livello fanno parte della formazione universitaria e costituiscono approfondimenti tematici autonomamente offerti dagli atenei. L'offerta è piuttosto varia; a titolo di esempio si possono indicare l'Executive Master in Data Management & Business Analytics presso l'Università IULM di Milano [66] e il Master in Business Intelligence e Big Data Analytics presso l'Università Bicocca di Milano [67]; sul sito web si possono trovare altri indirizzi di iniziative recenti.

MOOC

I MOOC (Massive Open Online Courses) sono dei corsi online aperti a tutti su larga scala, pensati per una formazione a distanza che coinvolga un numero elevato di utenti, provenienti da retroterra culturali molto differenziati e da diverse aree geografiche. Sono corsi di livello universitario, pensati però per un pubblico molto ampio. I corsi prevedono video di lezioni, materiale didattico, esercitazioni e forum di discussione. Offrono opportunità a chi studia ma anche al personale docente.

Federica.EU

Federica.EU [68] è la piattaforma MOOC dell'Università Federico II di Napoli; si presenta come "una fabbrica digitale per il futuro dell'istruzione universitaria". Sviluppata grazie al cofinanziamento dei fondi strutturali europei, con 150 corsi e oltre 100 milioni di learner, Federica è diventata una best practice nel panorama internazionale. Interessanti esempi sono il corso su Industria 4.0 [69] e i corsi forniti dall'Università di Milano-Bicocca, che ha stretto una partnership con Federica Weblearning promuovendo corsi sui Big Data in formato MOOC [70].

Coursera

Coursera [71] è un'azienda statunitense che opera nel campo delle tecnologie didattiche, fondata da docenti di informatica dell'Università di Stanford. Coinvolge un centinaio di università ed enti operanti nel campo dell'istruzione superiore in tutto il mondo. Sebbene il completamento e le lezioni siano gratuiti e forniscano un attestato di frequenza, per ottenere certificati ufficiali è generalmente richiesto un pagamento per l'iscrizione a una piattaforma di verifica individuale e coprire i costi amministrativi.

Futurelearn

Futurelearn [72] è la social learning platform della Open University con più di sei milioni di iscritti, che offre corsi online, gratis o a pagamento, dalle migliori università e istituzioni internazionali.

Udemy

Udemy [73] offre più di 100.000 corsi online on demand di alta qualità, anche in lingua italiana, con tanta offerta nell'AI anche in argomenti molto specifici, per esempio il catalogo sull'AI [74] e i corsi [75] specifici sul reinforcement learning. Con la specifica applicazione si può studiare anche su smartphone e tablet. Offre interessanti opportunità per diventare docenti e guadagnare con la vendita del corso.

Riviste scientifiche

Una rivista scientifica ha una direzione e un comitato editoriale composto da persone riconosciute come esperte nel settore, meglio ancora se la selezione degli articoli viene svolta da un gruppo di persone esperte incaricate di accettare solo il meglio delle proposte. Su questo genere di riviste vengono pubblicati gli articoli prodotti dalla ricerca scientifica in accademia e aziende interessate.
Le riviste scientifiche sono caratterizzate anche dall'avere l'Impact Factor, un indice bibliometrico sviluppato dall'Institute for Scientific Information (ISI) nel 1961 e attualmente di proprietà dell'editore Thomson Reuters. Misura il numero medio di cita-

zioni ricevute, nell'anno di riferimento considerato, dagli articoli pubblicati da una rivista scientifica nei due anni precedenti: è pertanto un indicatore della performance dei periodici scientifici, che esprime l'impatto di una pubblicazione sulla comunità scientifica di riferimento. L'Impact Factor di una rivista non esprime in sé un valore, ma va considerato rispetto ai valori raggiunti dai periodici del medesimo ambito disciplinare (subject category nel *Journal Citation Reports*), poiché ogni comunità è caratterizzata da un comportamento citazionale specifico. Su [76] si possono trovare altri dettagli sull'Impact Factor e link alle graduatorie delle riviste.
Difficilmente queste risorse offrono grandi listati di codice pronti all'uso, ma possono fornire qualche idea, sperimentazione, caso di studio, metodologia, nuovi algoritmi. Se si trova del codice utile, è meglio verificarne la proprietà intellettuale; molto spesso viene concesso codice libero solo per scopi di ricerca e non per l'uso commerciale.
Per trovare un articolo si può partire da un motore generico come Google, ma è meglio usare il servizio specifico Scholar oppure i motori più specifici offerti dagli editori come IEEE [77], Springer [78], ScienceDirect [79].
Le riviste scientifiche sono a pagamento, con libero accesso da università, centri di ricerca, biblioteche che le hanno acquistate.
Una risorsa gratuita molto utile è arXiv [80], un archivio per bozze definitive di articoli scientifici accessibile gratis tramite internet. In molti campi della scienza, la maggior parte delle pubblicazioni scientifiche è collocata nell'archivio arXiv. Google fornisce i link verso arXiv, oppure si può usare la ricerca interna; altri consigli per cercarli si possono trovare su [81].
Per avere qualche chiarimento, o per avere il codice, si può scrivere all'email dell'autore. È buona pratica e segno di rispetto citare l'autore nella bibliografia, se si usano i suoi lavori.

Siti web

La quantità di siti web sull'AI è ormai sterminata. Per non perdersi, è importante stabilire il tipo di fonte (università, azienda, professionista, consulente ecc.) che interessa e poi cercare i relativi siti web.
Giusto per rendere l'idea, qualche esempio in lingua italiana per conoscere l'AI riguarda le pagine dei siti [82] [83] [84] [85] [86] [87] e un utile glossario [88] per conoscere il significato dei principali termini in lingua inglese; si segnalano inoltre i siti web più specifici sulla programmazione, come Open Data Science [89], Kdnuggets [90] e AItrend [91].
Hackr.it si presenta su [92] con lo slogan "Trova il miglior corso e tutorial".
Da considerare anche Fluel.it, perché fornisce eventi formativi di taglio manageriale per mettere in contatto imprenditori e manager con i migliori rappresentanti delle tecnologie più disruptive.

Social Media

I social media offrono profili personali di persone esperte, gruppi di discussione, pagine associate a libri oppure ad aziende, opportunità di lavoro ecc. Una piazza digitale da frequentare, con i suoi vantaggi e svantaggi. Basta mettere un po' di MiPiace a post dedicati all'AI per essere classificati dalle piattaforme come persone interessate a certi argomenti, e in questo modo si riceveranno ulteriori post, pubblicità e consigli per altre pagine inerenti a queste tematiche.
Per trovare le risorse si può usare il motore di ricerca interno alla piattaforma preferita. Qualche esempio lo si può trovare in Facebook, come "Machine Learning Memes for Convolutional Teens" [93], con molti simpatici meme formati da immagine e testo, "Tarallucci, Vino e Machine Learning" [94], le associazioni AIxIA [95] e IAML [96]. Su LinkedIn si può trovare "Python for Data Science and Machine Learning" [97] e un altro gruppo su varie tematiche di programmazione [98].
Da segnalare Researchgate [99], un social network specifico per la ricerca scientifica, dove gli autori possono mettere a disposizione i propri articoli, ed è anche possibile richiederli direttamente. È necessario iscriversi, ma l'iscrizione è gratuita, ed è possibile seguire specifici autori o tematiche di interesse, spedire o ricevere messaggi e proposte di lavoro.

Video tutorial

Su YouTube si trovano numerosi video utili per la formazione, grazie a corsi universitari o creati da appassionati e a riprese fatte durante gli eventi.
Da segnalare la galleria di video della community ODSC [2.88].

Riferimenti bibliografici

A. Pipinellis, *GitHub, Piccolo manuale per lo sviluppo collaborativo di software*, Apogeo, Milano 2019.

Note

1 Associazione Italiana per l'Intelligenza Artificiale, https://aixia.it/
2 Associazione CVPL, https://www.cvpl.it/
3 Associazione IAML, https://www.iaml.it/
4 Associazione SIREN, https://www.siren.polito.it/
5 Blog di Jason Brownlee, https://machinelearningmastery.com/blog/
6 Blog Data science dojo, https://blog.datasciencedojo.com
7 Blog curato da IBM Data and AI, https://medium.com/inside-machine-learning
8 Towards data science, https://towardsdatascience.com/
9 kdnuggets, https://www.kdnuggets.com/news/index.html
10 Data Science Central, http://www.datasciencecentral.com/
11 *Top 75 Artificial Intelligence Websites and Blogs for AI Enthusiast*, https://blog.feedspot.com/ai_blogs/
12 *Top Active Blogs on AI, Analytics, Big Data, Data Science, Machine Learning*, https://www.kdnuggets.com/2019/01/active-blogs-ai-analytics-data-science.html
13 *5 Blogs and bloggers to follow if AI interests you (& one awwwwwsome professor!)*, https://medium.com/productivity-revolution/our-favourite-five-ai-bloggers-fd-6d0357601a
14 *40 Must-read AI Machine Learning blogs*, https://www.springboard.com/blog/machine-learning-blog/
15 *The Best AI and Machine Learning Blogs to Follow Religiously*, https://lionbridge.ai/articles/best-20-ai-and-machine-learning-blogs-to-follow-religiously/
16 Consorzio Interuniversitario Nazionale per l'Informatica, https://www.consorzio-cini.it/
17 Istituto Italiano di Tecnologia, http://www.iit.it
18 Laboratorio Computer Vision & Multimedia Università di Pavia, https://vision.unipv.it/
19 Git, https://git-scm.com/ e https://github.com/git
20 GitHub, https://github.com/
21 *GIT in pochi passi*, https://www.html.it/articoli/git-in-pochi-passi/
22 *Guida completa a GIT*, https://www.mrwebmaster.it/programmazione/guida-git/
23 *Come si usa GitHub*, https://github.com/emergenzeHack/terremotocentro/wiki/002-Come-si-usa-Github
24 Fluel, http://fluel.it/
25 AI Italia, http://aiitalia.it/
26 Telegram AI Italia, @aiitalia t.me/aiitalia
27 Gruppo Facebook AI Intelligenza Artificiale Italia, https://www.facebook.com/groups/1108748499235514/
28 Gruppo LinkedIn AI Intelligenza Artificiale Italia, https://www.linkedin.com/groups/8622106/

29 Data Science Milan, http://datasciencemilan.x10host.com/ e https://datasciencemilan.slack.com/
30 Pagina Facebook di Data Science Milan, https://www.facebook.com/DataScienceMilan
31 Gruppo LinkedIn Data Science Milan, https://www.linkedin.com/groups/8497363/
32 Data Science Milan, pagina su LinkedIn, https://www.linkedin.com/company/27062920
33 DataScienceSeed, http://www.datascienceseed.com/
34 Pagina Facebook di Deep Learning Italia, https://www.facebook.com/Deep-Learning-Italia-2111439285760266/ e gruppo https://www.facebook.com/groups/196584677432705/
35 Deep Learning Italia su LinkedIn, https://www.linkedin.com/company/deep-learning-italia/
36 Gruppo Meetup Deep Learning Italia, https://www.meetup.com/it-IT/Deep-Learning-Italia-Meetup-Group/
37 Kaggle, https://www.kaggle.com
38 *Kaggle: The Home of Data Science*, http://www.mathisintheair.org/wp/2016/11/kaggle-the-home-of-data-science/
39 ODSC, https://odsc.com/
40 Gruppo Meetup ODSC Milano Data Science, https://www.meetup.com/it-IT/Milano-Data-Science-ODSC/
41 ODSC formazione, https://learnai.odsc.com/
42 StackOverFlow, https://stackoverflow.com/
43 *Stack Overflow Annual Developer Survey*, https://insights.stackoverflow.com/survey
44 Stack Overflow sezione Jobs, https://stackoverflow.com/jobs
45 *Python è sempre tra i linguaggi più ricercati su StackOverflow*, https://www.html.it/17/09/2019/python-e-sempre-tra-i-linguaggi-piu-ricercati-su-stackoverflow/
46 The Data Literacy Project, https://thedataliteracyproject.org/
47 *Usare Meetup*, https://help.meetup.com/hc/it/categories/360000087852-Usare-Meetup
48 Pagina italiana di Meetup, https://www.meetup.com/it-IT/
49 *Evento Meetup #10 MLMilan: Deep dive into Reinforcement Learning - Use case in Finance*, https://www.meetup.com/it-IT/Machine-Learning-Milan/events/264428176/
50 Profilo Meetup di Roberto Marmo, https://www.meetup.com/members/198551754/?_locale=it-IT
51 Gruppi Meetup di Scienza dei dati attivi in Italia, https://www.meetup.com/it-IT/topics/data-science/it/

52 Gruppi Meetup di intelligenza artificiale attivi in Italia, https://www.meetup.com/it-IT/topics/ai/it/
53 Gruppi Meetup di machine learning attivi in Italia, https://www.meetup.com/it-IT/topics/machine-learning/it/
54 Gruppo Meetup Data Science Milan, https://www.meetup.com/it-IT/Data-Science-Milan/
55 Gruppo Meetup Machine Learning Milan, https://www.meetup.com/it-IT/Machine-Learning-Milan/
56 Gruppo Meetup Milan Women in Machine Learning and Data Science, https://www.meetup.com/it-IT/Milan-Women-in-Machine-Learning-and-Data-Science/
57 Gruppo Meetup Workshop in Data Science, https://www.meetup.com/it-IT/Workshop-in-Data-Science/
58 Gruppo Data Science Meetup Pavia, https://www.meetup.com/it-IT/Data-Science-Meetup-Pavia/Data Science Milan
59 Eventbrite suggerisce eventi di machine learning a Milano, https://www.eventbrite.it/d/italy--milano/machine-learning/
60 Eventbrite suggerisce eventi di intelligenza artificiale a Milano, https://www.eventbrite.it/d/italy--milano/artificial-intelligence/
61 DSC Data Science Search Engine cheat sheet, https://www.datasciencecentral.com/page/search?q=cheat+sheet
62 List of Data Science Cheatsheets to rule the world, https://github.com/FavioVazquez/ds-cheatsheets
63 Infografiche riguardanti intelligenza artificiale in Pinterest.it, https://www.pinterest.it/robertomarmo/intelligenza-artificiale/
64 OPAC SBN catalogo del servizio bibliotecario nazionale, https://opac.sbn.it/opacsbn/opac/iccu/free.jsp
65 Google Ricerca Libri, https://books.google.it
66 Executive Master in Data Management & Business Analytics presso Università IULM di Milano, https:www.masterdmba.it
67 Master in Business Intelligence e Big Data Analytics presso Università Bicocca di Milano, https://www.unimib.it/didattica/master-universitari
68 Federica.eu, https://www.federica.eu/tutti_i_mooc
69 Corso su Industria 4.0 in Federica.eu, https://www.federica.eu/c/industria_40
70 I corsi di Data Science di Milano-Bicocca sono online con Federica.EU, https://www.federica.eu/blog/2019/03/19/data-science-milano-bicocca-federica/ e https://www.federica.eu/bicocca/
71 Coursera, https://www.coursera.org
72 Futurelearn, https://www.futurelearn.com/
73 Udemy, https://www.udemy.com/

74 Udemy corsi su intelligenza artificiale, https://www.udemy.com/topic/artificial-intelligence/?persist_locale&locale=it_IT
75 Udemy corsi su reinforcement learning, https://www.udemy.com/topic/reinforcement-learning/
76 Impact Factor (IF), http://biblioteche.unipv.it/home/risorse/indicatori-bibliometrici/impact-factor-if
77 Motore di ricerca dell'editore scientifico IEEE, https://ieeexplore.ieee.org/Xplore/home.jsp
78 Motore di ricerca dell'editore scientifico Springer, https://www.springer.com/east/home?SGWID=5-102-13-0-0
79 Motore di ricerca nel database ScienceDirect, https://www.sciencedirect.com/
80 arXiv, https://arxiv.org/
81 *Dove trovare gli articoli scientifici su internet*, https://blog.sitd.it/2017/04/06/trovare-gli-articoli-scientifici-internet/
82 Intelligenza artificiale, http://www.intelligenzaartificiale.it/
83 ZeroUno, https://www.zerounoweb.it/osservatori/ai-e-cognitive-computing/
84 AI4business, https://www.ai4business.it/
85 *Corriere Comunicazioni*, https://www.corrierecomunicazioni.it/tag/intelligenza-artificiale/
86 *Il Sole-24 Ore*, https://nova.ilsole24ore.com/argomento/intelligenza-artificiale/
87 key4biz, https://www.key4biz.it/tag-2/intelligenza-artificiale/
88 Glossario Accenture su applied intelligence, https://www.accenture.com/it-it/applied-intelligence-glossary
89 Open Data Science, https://opendatascience.com
90 Kdnuggets, https://www.kdnuggets.com
91 AItrend, https://www.aitrends.com/
92 Find the Best Data Science Courses & Tutorials, https://hackr.io/data-science
93 Machine Learning Memes for Convolutional Teens, https://www.facebook.com/convolutionalmemes/
94 Tarallucci, Vino e Machine Learning, https://www.facebook.com/groups/TarallucciVinoMachineLearning/
95 Associazione Italiana per l'Intelligenza Artificiale, https://www.facebook.com/AIxIA1988/
96 Italian Association for Machine Learning, https://www.facebook.com/machinelearningitalia/
97 Python for Data Science and Machine Learning, https://www.linkedin.com/groups/8288265/
98 Computer Vision, Deep Learning, Deep Reinforcement Learning, OpenCV, C++, Caffe, TensorFlow, GANs, https://www.linkedin.com/groups/10320678/
99 Researchgate, https://www.researchgate.net

APPENDICE D
USARE RAPIDMINER

a cura di Rodolfo Baggio, Filippo Carone Fabiani

In questa appendice viene spiegato come usare RapidMiner, una piattaforma di data mining e predictive analytics che permette di realizzare modelli di AI in maniera visuale anche senza possedere particolari conoscenze di programmazione. In tal modo, si può subito cominciare a sperimentare qualcosa.
I paragrafi successivi contengono esempi riguardanti email spam, analisi degli acquisti in un supermercato, sentiment analysis dei testi, previsione del trend delle vendite.
Chi non è attratto dall'uso di questo software può comunque trovare di interesse i dataset considerati e le metodologie di analisi.

Gli autori

Rodolfo Baggio è docente e coordinatore dell'area di Strategie Digitali al Master in Economia del turismo e Research Fellow del Centro Dondena per la Ricerca sulle dinamiche sociali e politiche pubbliche dell'Università Bocconi di Milano, profilo LinkedIn su [1].
Filippo Carone Fabiani si laurea in Fisica teorica alla Sapienza di Roma lavorando su problemi di meccanica statistica quantistica e consegue il Dottorato di ricerca presso l'Università Bicocca di Milano su problemi di trasporto quantistico. Dopo un Master in Sistemi complessi, conseguito allo IUSS di Pavia, inizia a occuparsi di Machine Learning e Intelligenza Artificiale, occupando la posizione di ricercatore nel settore accademico e privato; è assegnista di ricerca presso l'Università di Bergamo, responsabile di un progetto sulla classificazione di segnali elettroencefalici legati alle attività linguistiche e motorie. Profilo LinkedIn su [2].

Che cos'è RapidMiner

RapidMiner è una piattaforma software open source per fare prototipazione e sviluppo rapidi di applicazioni di analisi dei dati. Include un ambiente integrato per la preparazione dei dati, estrazione del testo, analisi predittiva, modelli di machine learning. La forza particolare di RapidMiner risiede nell'analisi predittiva, quindi nella previsione degli sviluppi futuri sulla base dei dati raccolti. Scritto in Java, offre un'interfaccia grafica in lingua inglese, utilizzabile facilmente senza particolari conoscenze di programmazione tramite le operazioni del tipo drag and drop (successione di tre azioni: cliccare su un oggetto, trascinarlo in un'altra posizione, rilasciarlo), come mostrato in Figura D.1.

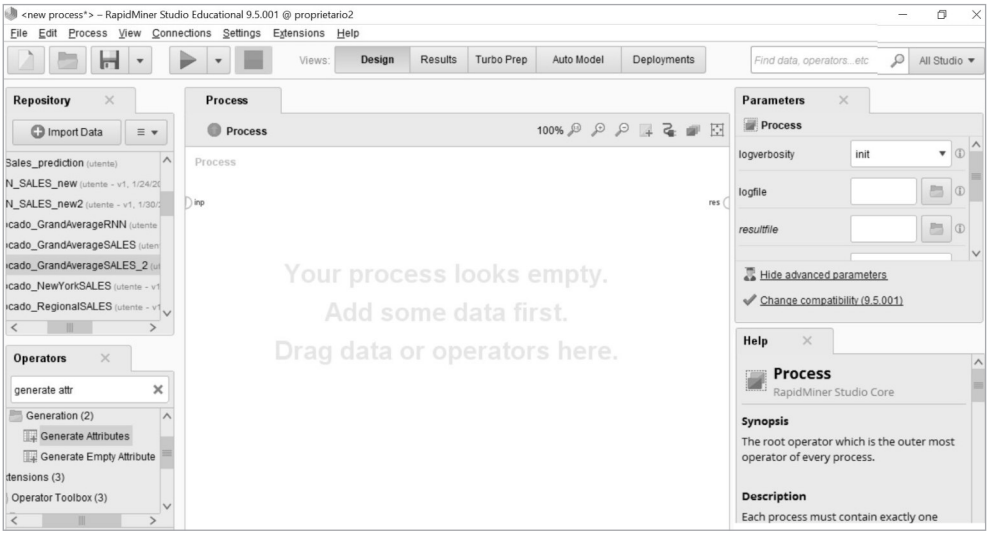


Figura D.1 – Interfaccia grafica all'apertura di RapidMiner.

Dataset e operatori si presentano come blocchi rettangolari che vengono inseriti direttamente nel workspace centrale (area Design), trascinandoli dai tab laterali (aree Repository e Operators), una volta individuati attraverso la finestra di ricerca relativa. Tutti gli oltre 500 operatori sono configurabili selezionandoli con un doppio clic e modificando i campi del tab Parameters corrispondente che si apre a destra dell'area Design. Nel seguito del testo, ogni operatore è identificabile con la lettera iniziale maiuscola.

Importa tabelle Excel, file SPSS e record di molti database, integra strumenti di data mining come WEKA e R, offre varie possibilità di visualizzazione dei risultati, e può essere esteso con plugin aggiuntivi disponibili nel RapidMiner Marketplace.

RapidMiner supporta tutti i passaggi del processo di data mining con tre moduli: RapidMiner Studio, RapidMiner Server e RapidMiner Radoop.

I flussi di lavoro sono chiamati processi (process) e sono costituiti da operatori collegati in cascata connessi attraverso specifiche porte, così ogni operatore esegue una singola attività all'interno del processo e il suo output diventa l'input per il successivo operatore. RapidMiner può essere richiamato da altri programmi e utilizzato come API.

Dove scaricare

Il sito di riferimento è [3], i prezzi delle licenze si trovano su [4], la versione di prova gratis per 30 giorni si scarica da [5], studenti e professori possono scaricare gratis da [6] una licenza annuale rinnovabile e hanno accesso a vari servizi formativi.

Nel sito web https://www.algoritmiia.it/ associato al libro *Algoritmi per l'intelligenza artificiale* si possono scaricare i file con i processi esportati per poterli subito eseguire.

Vantaggi rispetto a Python

Con RapidMiner si può programmare un'analisi dati semplicemente tracciando rettangoli e linee, come in un disegno, mentre con Python occorre scrivere linee di codice con le dovute difficoltà, però la programmazione permette di creare soluzioni su misura e offre maggiore flessibilità. Un'ampia tabella con vari confronti si trova su [7].
Si potrebbe fare una prima versione con RapidMiner, per sviluppare subito un prototipo da approfondire successivamente con Python.
Volendo, questi due contesti si possono integrare come spiegato in [8].

Letture consigliate

Su [9] sono disponibili due importanti manuali in lingua inglese, pubblicati nella specifica pagina della documentazione su [10]. Due utili libri sono stati scritti da Chisholm e Hofmann.

Email spam

Il primo esempio riguarda la costruzione di un modello di classificazione di messaggi email per identificare possibili spam. Il dataset usato è reperibile sul sito dell'UCI Machine Learning Repository [11]; esso contiene 4600 messaggi, per ognuno dei quali vengono forniti 57 valori di vario tipo, dalla presenza di certe parole alla lunghezza delle frasi, a quante parole sono scritte in caratteri maiuscoli ecc. Ogni messaggio è stato classificato manualmente come spam o no.
Il dataset in formato CSV va caricato nel repository locale di RapidMiner. Durante l'importazione eseguita dal menu File data si possono ridefinire formati e funzioni delle variabili, come nella Figura D.2, in cui viene assegnato alla variabile *is spam* il ruolo label comunicando al software che questo è il campo usato per la classificazione.
Per costruire il modello e scegliere il migliore algoritmo per i dati disponibili viene usato l'operatore Cross Validation, che, in maniera automatica, divide il dataset in una parte di prova usata per addestrare l'algoritmo scelto e l'altra come test, effettuando questa operazione un certo numero di volte in modo da compensare possibili problemi dovuti alla scelta casuale dei record delle due parti. L'operatore è nidificato e contiene due fasi: nella prima l'algoritmo scelto viene addestrato sui dati di prova, nella seconda il modello viene applicato alla parte test e il risultato è confrontato con la classificazione esistente per calcolare il grado di affidabilità della soluzione scelta. La Figura D.3 mostra l'operatore di Cross Validation, con la tipica organizzazione grafica di RapidMiner composta da moduli rettangolari e linee che uniscono i punti di ingresso e di uscita formati da semicerchi.

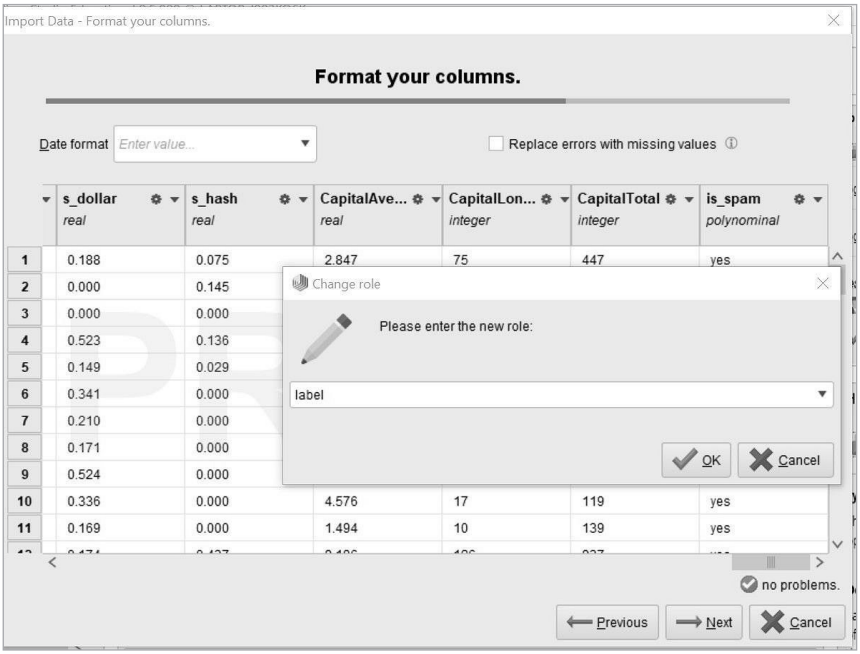


Figura D.2 – Caricamento del file CSV con dataset sullo spam e assegnazione dei ruoli nei campi.

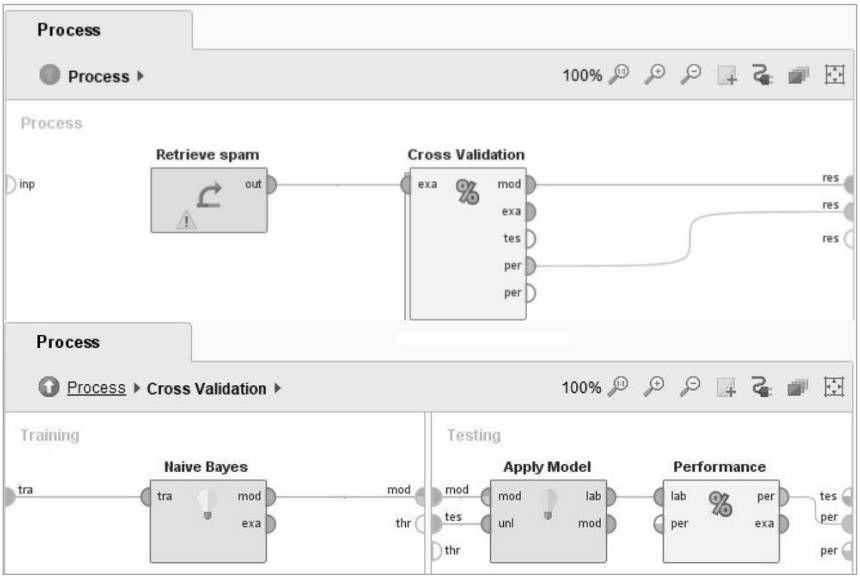


Figura D.3 – Processo di Cross Validation per organizzare training e test set. Si può notare l'approccio visual di RapidMiner composto da moduli rettangolari e linee di connessione.

Il risultato è la matrice di confusione contenente la misura dell'accuratezza dell'algoritmo scelto. A questo punto basta provare diversi algoritmi, come mostrato in Figura D.4, e scegliere quello che produce i risultati migliori. Qui, per esempio, vengono usati Naive Bayes e Support Vector Machine (SVM). Il secondo, come si nota, produce risultati più adeguati.

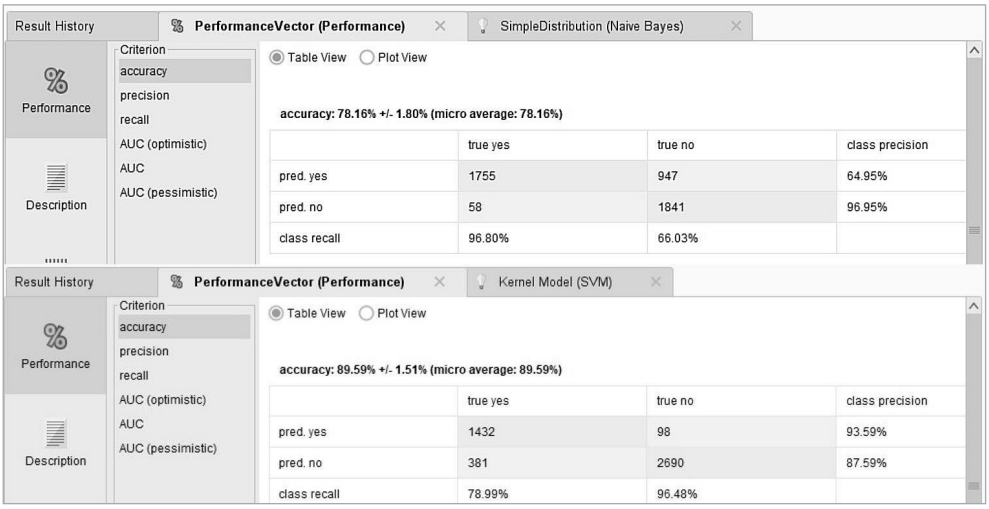


Figura D.4 – Matrici di confusione risultanti da prove con modelli diversi.

Una volta scelto l'algoritmo migliore, è possibile utilizzarlo per classificare dati nuovi, non classificati. Per fare ciò si riaddestra l'algoritmo con i dati classificati e lo si applica al nuovo dataset che ha, ovviamente, le stesse variabili a eccezione del campo *is spam*, come mostrato in Figura D.5.

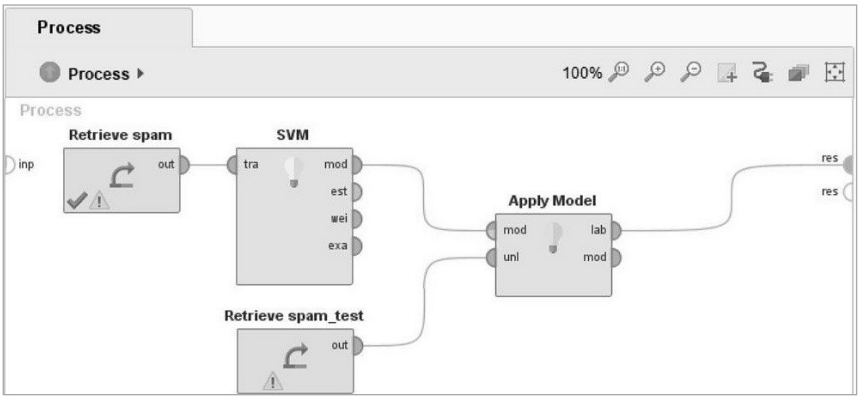


Figura D.5 – Processo di classificazione dei dati non coinvolti nel training set.

Il risultato è un nuovo dataset classificato che contiene anche la *confidence* delle due possibili classificazioni (yes/no), come mostrato in Figura D.6.

Row No.	prediction(is_spam)	confidence(yes)	confidence(no)	Make	Address	All	3D
1	no	0.066	0.934	0	0	0	0
2	no	0.024	0.976	0	0	0.510	0
3	no	0.148	0.852	0	0	0	0
4	no	0.029	0.971	0	0	0	0
5	no	0.102	0.898	0.100	0	0	0
6	no	0.150	0.850	0	0	0	0
7	yes	0.930	0.070	0	0	0.680	0
8	no	0.092	0.908	0	0	3.840	0
9	no	0.460	0.540	0.050	0.050	0.400	0
10	no	0.046	0.954	0	0	1.260	0

Figura D.6 – Risultati sui dati non coinvolti nel training set.

Regole di associazione

Le regole di associazione sono tecniche derivate dal data mining per scoprire regolarità all'interno di catene di transazioni, in particolare per estrarre relazioni nascoste tra i dati. Per esempio, analizzando gli scontrini di un supermercato si trova la regola { cipolle , patate } ⇒ { hamburger } utile per descrivere il comportamento di acquisto secondo cui, se il cliente compra insieme cipolle e patate, è probabile che acquisti anche la carne per hamburger.
Il dataset utilizzato è il Groceries Market Basket Dataset, distribuito su [12], formato da 9835 scontrini di clienti di un supermercato contenenti l'elenco dei prodotti acquistati; un esempio è riportato nella Figura D.7.

products
citrus_fruit,semi-finished_bread,margarine,ready_soups
tropical_fruit,yogurt,coffee
whole_milk
pip_fruit,yogurt,cream_cheese,meat_spreads
other_vegetables,whole_milk,condensed_milk,long_life_bakery_product
whole_milk,butter,yogurt,rice,abrasive_cleaner
rolls_buns
other_vegetables,UHT-milk,rolls_buns,bottled_beer,liquor_(appetizer)
pot_plants
whole_milk,cereals
tropical_fruit,other_vegetables,white_bread,bottled_water,chocolate
citrus_fruit,tropical_fruit,whole_milk,butter,curd,yogurt,flour,bottled_water
... ...

Figura D.7 – Prodotti nel Groceries Market Basket Dataset.

Le regole di associazione vengono ricavate dopo aver dedotto i frequent itemset, ovvero le combinazioni di prodotti acquistati che ricorrono più spesso nel dataset.
Dopo aver caricato i dati si procede alla loro preparazione per l'operatore FP-Growth che calcola i frequent itemset. Essenzialmente, tutti i diversi elementi (prodotti) contenuti nelle righe diventano variabili dicotomiche (Yes/No) per ogni riga (uno scontrino). Questi dati costituiscono l'input per il calcolo dei frequent itemset e delle regole di associazione tramite il processo indicato nella Figura D.8.

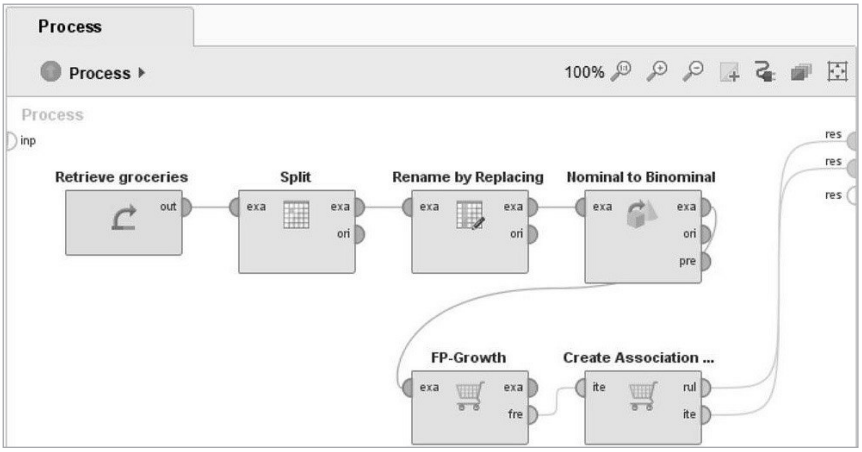


Figura D.8 – Processo per il calcolo dei frequent itemset e delle regole di associazione.

Come output si ricavano i frequent itemset, mostrati in Figura D.9, e le regole di associazione date in forma esplicita, mostrate in Figura D.10, o come tabella con le varie misure calcolate, mostrate in Figura D.11.

AssociationRules (Create Association Rules) | FrequentItemSets (FP-Growth)

No. of Sets: 45
Total Max. Size: 3
Min. Size: 2
Max. Size: 4
Contains Item:
Update View

Size	Support ↓	Item 1	Item 2	Item 3
2	0.075	whole_milk	other_vegetables	
2	0.057	whole_milk	rolls_buns	
2	0.056	whole_milk	yogurt	
2	0.049	whole_milk	root_vegetables	
2	0.047	other_vegetables	root_vegetables	
2	0.043	other_vegetables	rolls_buns	
2	0.043	other_vegetables	yogurt	
2	0.042	whole_milk	tropical_fruit	
2	0.040	whole_milk	soda	

Figura D.9 – Frequent itemset ottenuti dal processo mostrato in Figura D.8.

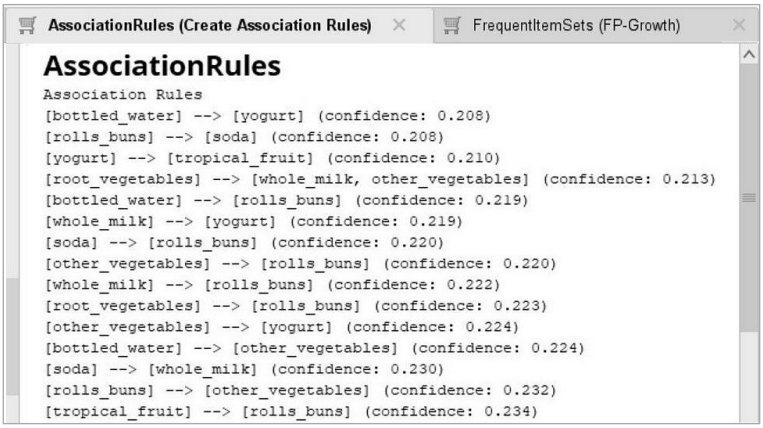
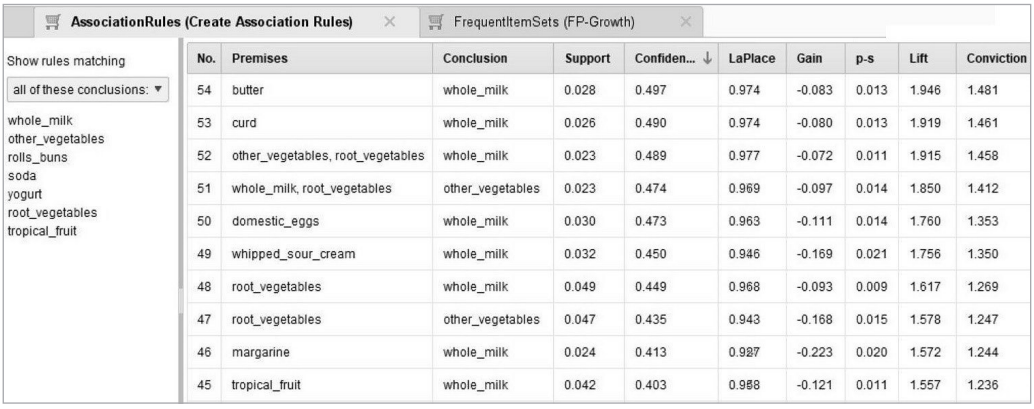


Figura D.10 – Regole di associazione date in forma esplicita.



No.	Premises	Conclusion	Support	Confiden...	LaPlace	Gain	p-s	Lift	Conviction
54	butter	whole_milk	0.028	0.497	0.974	-0.083	0.013	1.946	1.481
53	curd	whole_milk	0.026	0.490	0.974	-0.080	0.013	1.919	1.461
52	other_vegetables, root_vegetables	whole_milk	0.023	0.489	0.977	-0.072	0.011	1.915	1.458
51	whole_milk, root_vegetables	other_vegetables	0.023	0.474	0.969	-0.097	0.014	1.850	1.412
50	domestic_eggs	whole_milk	0.030	0.473	0.963	-0.111	0.014	1.760	1.353
49	whipped_sour_cream	whole_milk	0.032	0.450	0.946	-0.169	0.021	1.756	1.350
48	root_vegetables	whole_milk	0.049	0.449	0.968	-0.093	0.009	1.617	1.269
47	root_vegetables	other_vegetables	0.047	0.435	0.943	-0.168	0.015	1.578	1.247
46	margarine	whole_milk	0.024	0.413	0.927	-0.223	0.020	1.572	1.244
45	tropical_fruit	whole_milk	0.042	0.403	0.958	-0.121	0.011	1.557	1.236

Figura D.11 – Regole di associazione come tabella con le misure calcolate.

Anche in questo caso, come nel precedente, l'operatore Apply Association Rules consente di utilizzare le regole ottenute. L'operatore Subprocess è un contenitore in cui sono inserite le funzioni viste prima per la preparazione dei dati, come mostrato nella Figura D.12.

I risultati danno nuovi accoppiamenti possibili con i valori di confidence mostrati nella Figura D.13, con cui prevedere possibili accoppiamenti o costruire un prototipo per creare il sistema di raccomandazione dei prodotti da acquistare secondo il comportamento del cliente.

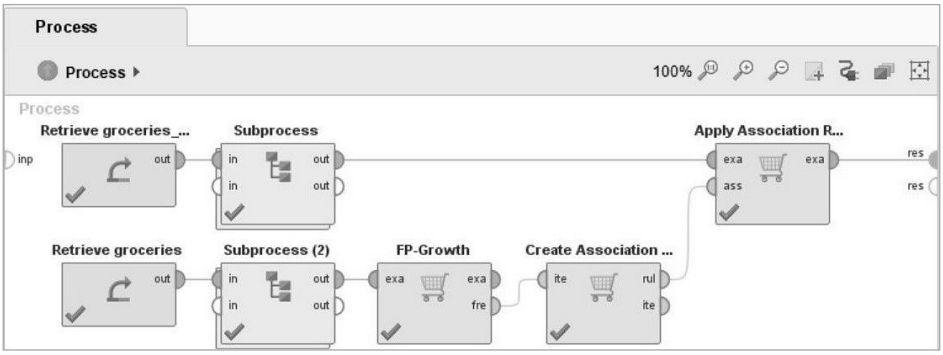


Figura D.12 – Operatore Subprocess per contenere le funzioni di preparazione dei dati.

	prodotti scelti				
	1	2	3	4	5
possibili altri prodotti (confidence)	citrus_fruit,margarine	yogurt,butter	coffee,sugar	whole_milk, berries	flour,salt,bottled_water
bottled_water		0.165		0.135	
butter				0.108	
citrus_fruit				0.119	
curd				0.102	
domestic_eggs				0.117	
fruit_vegetable_juice				0.104	
newspapers				0.107	
other_vegetables	0.349	0.311		0.293	0.224
pastry				0.130	
pip_fruit				0.118	
rolls_buns		0.246		0.222	0.219
root_vegetables		0.185		0.191	
sausage				0.117	
shopping_bags					
soda		0.196		0.157	0.262
tropical_fruit		0.210		0.166	
whipped_sour_cream				0.126	
whole_milk	0.413	0.497			0.311
yogurt				0.219	0.208

Figura D.13 – Accoppiamenti possibili con i valori di confidence mostrati.

Sentiment analysis nei testi

Per analizzare le emozioni scritte nei testi bisogna usare l'estensione denominata Text Processing da installare a parte con [13].

Il dataset utilizzato si chiama Hotel Reviews; scaricabile su [14], contiene 10.000 recensioni di hotel con la valutazione espressa su una scala da 1 (peggiore) a 5. Valutazioni minori di 2,5 sono considerate con sentiment negativo, le altre positive. Il modello di classificazione, da utilizzare poi eventualmente per analizzare recensioni non classificate, viene validato attraverso una Cross Validation come visto precedentemente.

Prima di fare ciò bisogna usare le funzioni di Text Processing per la preparazione dei testi con i seguenti passaggi:

1. il formato viene ridefinito come "text";
2. il testo viene diviso in parole (tokenization);
3. si trasforma tutto in caratteri minuscoli per facilitare confronti;
4. si ignorano le parole con meno di 3 caratteri (tipicamente non significative);
5. si eliminano le parole comuni come congiunzioni, articoli ecc. (stopwords);
6. si convertono verbi, plurali ecc. alla loro forma base (stemming).

Queste funzioni vengono raccolte nel contenitore Process Documents from Data mostrato in Figura D.14.

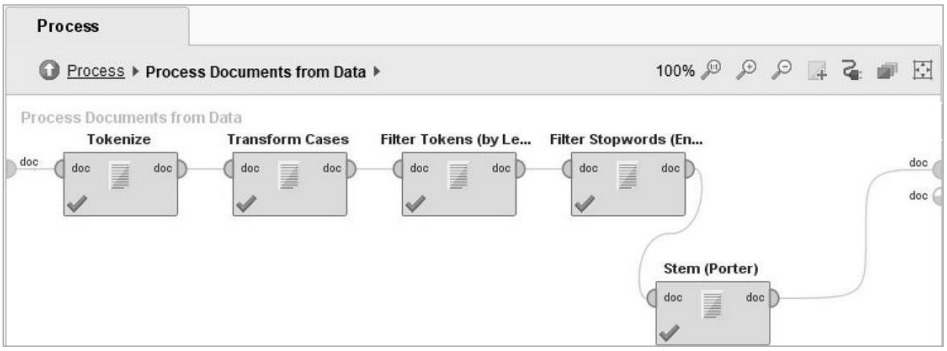


Figura D.14 – Contenitore Process Documents from Data per analisi del testo.

Il resto della catena di procedura è simile a quanto già fatto per il processo di classificazione, come mostrato nella Figura D.15.

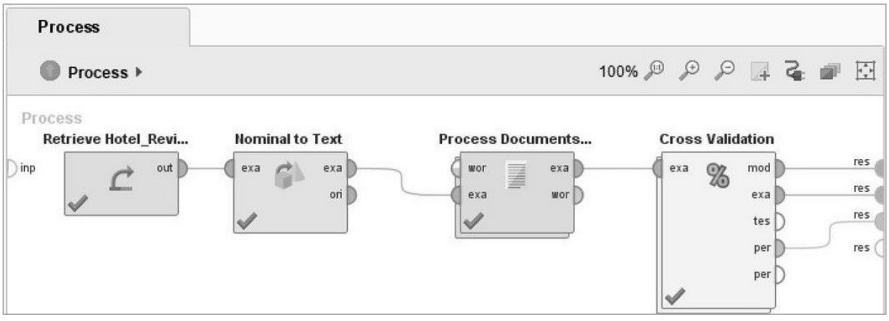


Figura D.15 – Processo di classificazione per analisi del testo.

Come mostrato in Figura D.16, viene usato un classificatore SVM; è possibile provare diversi algoritmi e scegliere il più adatto.

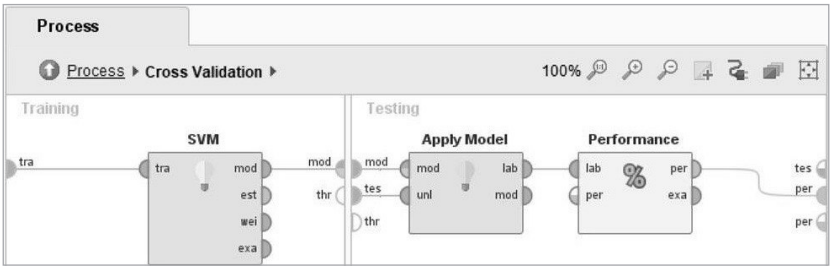
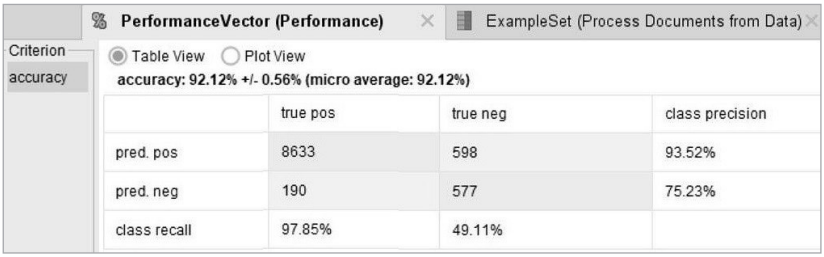


Figura D.16 – Processo di classificazione per analisi del testo con SVM.

Il risultato è più che soddisfacente, come mostra la matrice di confusione in Figura D.17



	true pos	true neg	class precision
pred. pos	8633	598	93.52%
pred. neg	190	577	75.23%
class recall	97.85%	49.11%	

Figura D.17 – Matrice di confusione dall'analisi del testo con SVM.

Come per i casi precedenti, una volta trovato l'algoritmo più efficace si può procedere a classificare testi non ancora valutati, come mostrato in Figura D.18. L'unica avvertenza, in questo caso, consiste nel connettere i due Process Documents in modo che venga formato un dizionario comune di parole, altrimenti c'è la possibilità che i due insiemi siano diversi e che l'algoritmo di classificazione non riesca a produrre un risultato adeguato.

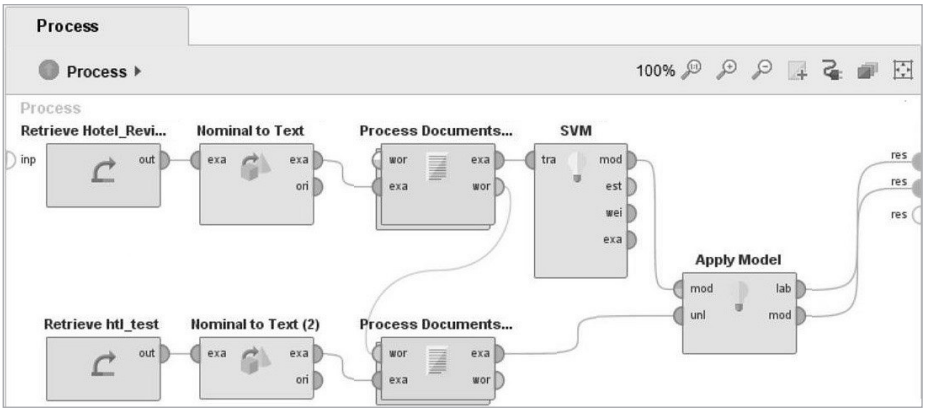


Figura D.18 – Processo per classificare testi non ancora valutati.

Previsione del trend delle vendite

In questo esempio viene usata una particolare neural network con addestramento supervisionato per prevedere il prezzo medio dell'avocado negli Stati Uniti in base al trend delle vendite. Viene seguito l'approccio standard per la creazione di un processo per fare AI.
Il dataset in formato CSV si può prendere dalla raccolta Kaggle su [15]; contiene la serie storica tra il 2015 e il 2018 dei prezzi e dei volumi di vendita di avocado in differenti mercati negli Stati Uniti. È formato da 18 attributi, di cui vengono scelti solo i seguenti, perché gli altri sono ridondanti o non determinanti per la previsione:

1. Att1: numero di settimane, numero intero nell'intervallo [0-52];
2. AveragePrice: prezzo medio unitario dell'avocado rilevato in dollari;
3. TotalaValume: volume totale di avocado scambiato;
4. Region: area geografica del distributore su 54 aree geografiche;
5. Type: tipologia di avocado organic e conventional.

Ogni riga corrisponde a un determinato rivenditore con cadenza settimanale nell'arco degli anni 2015-2018. Att1 è l'etichetta corrispondente alla settimana considerata e, per ciascun anno, va dal valore 0, corrispondente a "Mar 25, Year", al valore 51, corrispondente a "Jan 4, Year". I dati devono essere scaricati dal sito e caricati su un repository di RapidMiner.

Import di dataset

Con un clic su Import Data si può caricare il dataset avocado.csv come già fatto per la Figura D.1, quindi trascinarlo nella finestra Design per iniziare il disegno del modello. RapidMiner indica la tabella dei valori del dataset caricato con Retrieve nome-file; nel Design appare il blocco Retrieve avocado come mostrato in Figura D.19.

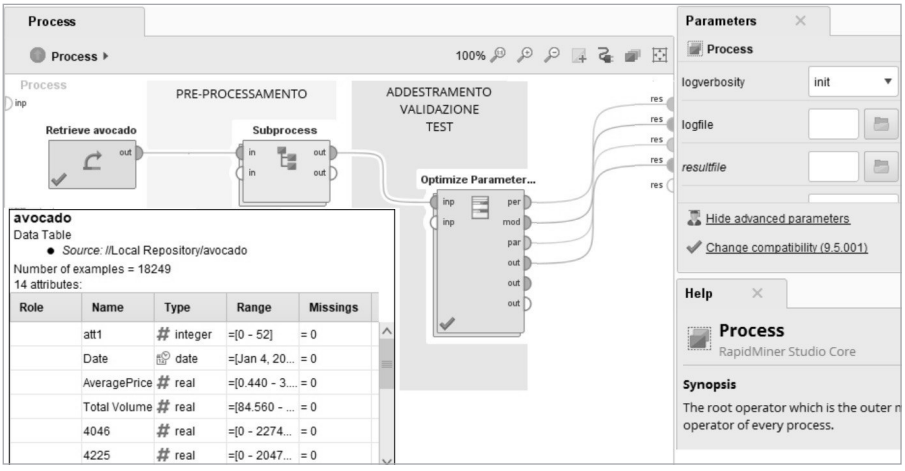


Figura D.19 – Processo principale applicato ai dati avocado.csv.

Pre-processamento dei dati

Si passa al pre-processamento dei dati, consistente in una serie di operazioni collegate in serie. Per comodità e chiarezza grafica del processo, le operazioni di pre-processamento vengono raggruppate in un unico blocco con Subprocess nel tab Operators e connettendo la porta out di Retrieve avocado alla porta di Subprocess, come mostrato in Figura D.20.

Prima di tutto, selezionare solo i quattro attributi rilevanti del problema, oltre all'attributo target: AveragePrice, Total Volume, Region e Type. Le operazioni da svolgere riguardano SelectAttributes, poi Parameters; scegliere il tipo di filtro con il campo Attribute filter type e creare la lista degli attributi utilizzati con il campo Attributes, come in Figura D.21.

Con un doppio clic all'interno di Subprocess, si entra nell'area di Design del sottoprocesso dove inserire tutti gli operatori coinvolti collegandoli in serie.

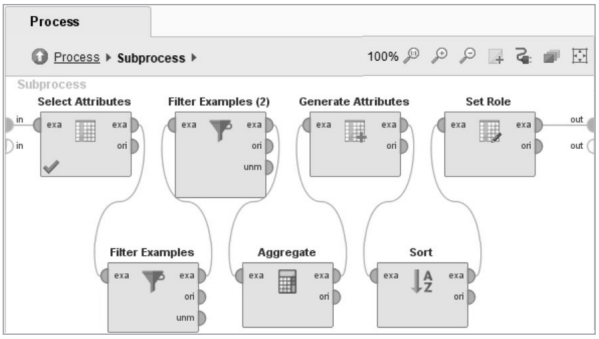


Figura D.20 – Sottoprocesso contenente tutti gli operatori di pre-processing.

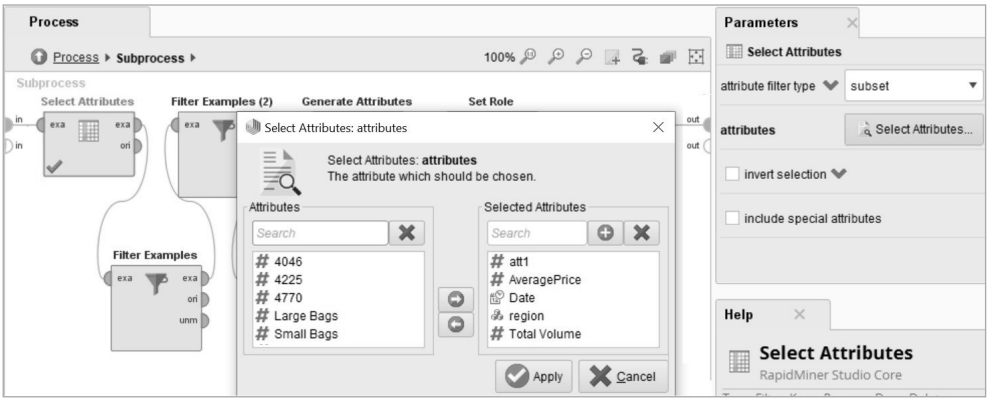


Figura D.21 – SelectAttribute, per la selezione degli attributi rilevanti.

Per avere dati affidabili e per non appesantire la trattazione con le problematiche relative al trattamento dei dati mancanti, bisogna filtrare tutte le occorrenze con missing values; in questo caso non ci sono missing values. Le operazioni da svolgere riguardano FilterExample, poi Parameters; scegliere il tipo di filtro con il campo Condition class, impostare l'elenco di condizioni con il campo Filters, spuntare il campo Match any, come mostrato in Figura D.22.

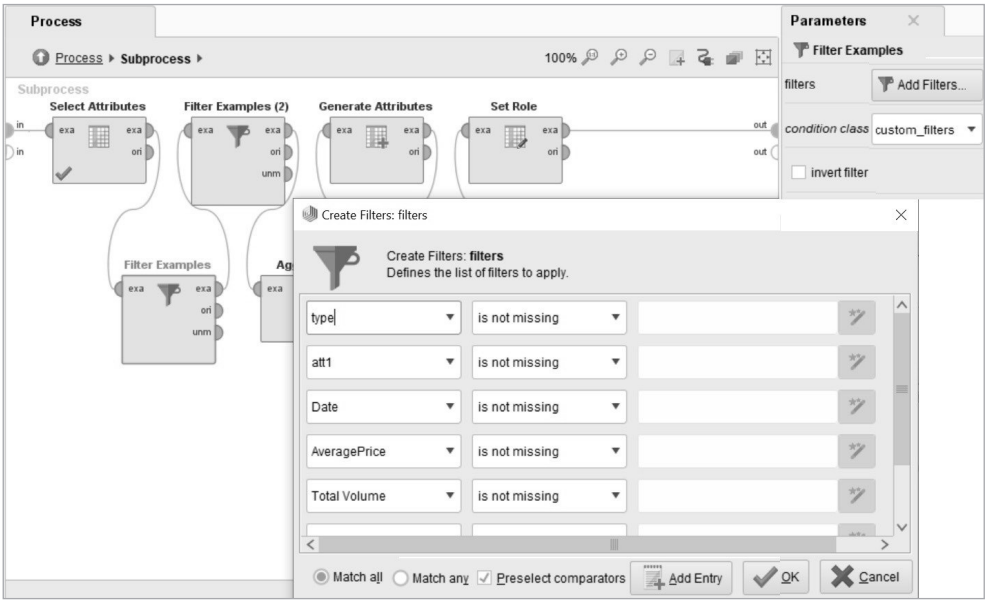


Figura D.22 – Utilizzo di FilterExample per escludere le occorrenze con missing values.

Il dataset riporta per ciascun anno di rilevazione e per ciascuna area i valori dei prezzi per le tipologie organic e conventional. Come si vede dalla Figura D.23, gli andamenti delle due tipologie sono sensibilmente differenti; meglio separare i dati secondo queste due classi e svolgere l'analisi per entrambe.
Per semplicità, viene mostrato il processo solo per il tipo organic; per quello conventional basta modificare il valore dell'attributo nel filtro e mantenere il resto inalterato. Le operazioni da svolgere riguardano FilterExample, poi Parameters; scegliere il tipo di filtro con il campo Condition class e Custom filter, impostare le condizioni con i campi in Filters e impostare il valore organic, come mostrato in Figura D.24.

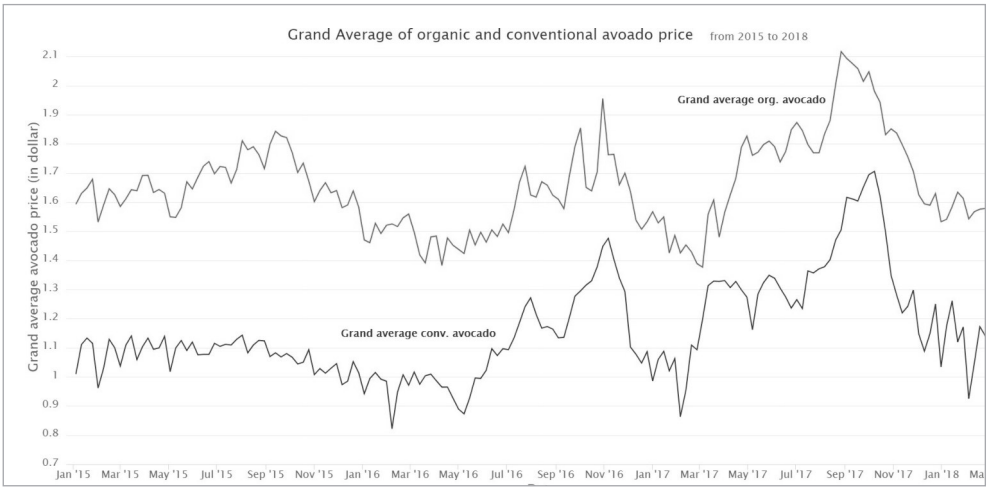


Figura D.23 – Media aggregata di tutti i distributori. In alto la media relativa al tipo organic e in basso, ben distinta, quella relativa al tipo conventional.

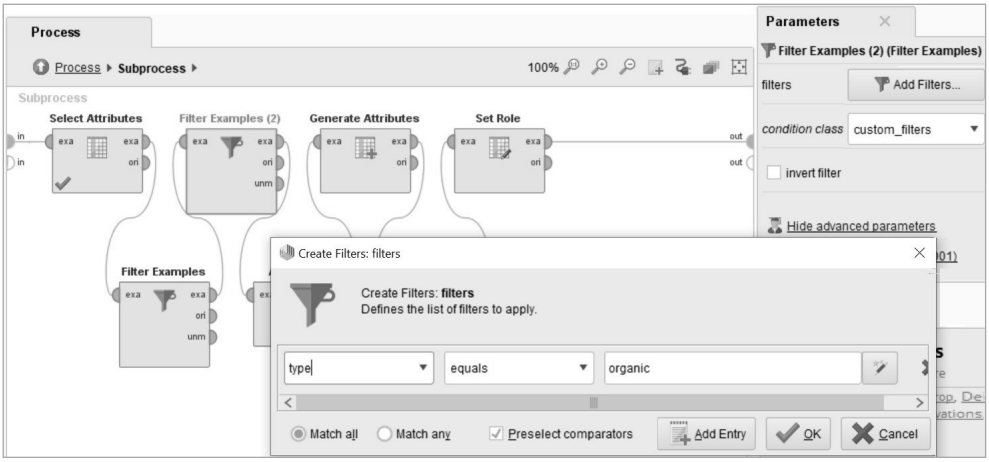


Figura D.24 – FilterExample per selezionare solo il tipo organic.

Poiché bisogna calcolare il valore medio sul numero dei rivenditori della variabile AveragePrice, bisogna prima raggruppare le occorrenze per data, utilizzando il tab Grouped by attributes con valore Date, attribuire alle variabili AveragePrice e TotalVolume la somma dei valori attraverso il tab Aggregation attribute. Le operazioni da svolgere riguardano Aggregate, poi Parameters; raggruppare per data con il campo Grouped by attributes con valore Date, assegnare i valori aggregati con il campo Aggregation attributes e Aggregation function con valore sum, come mostrato nelle Figure D.25 e D.26.

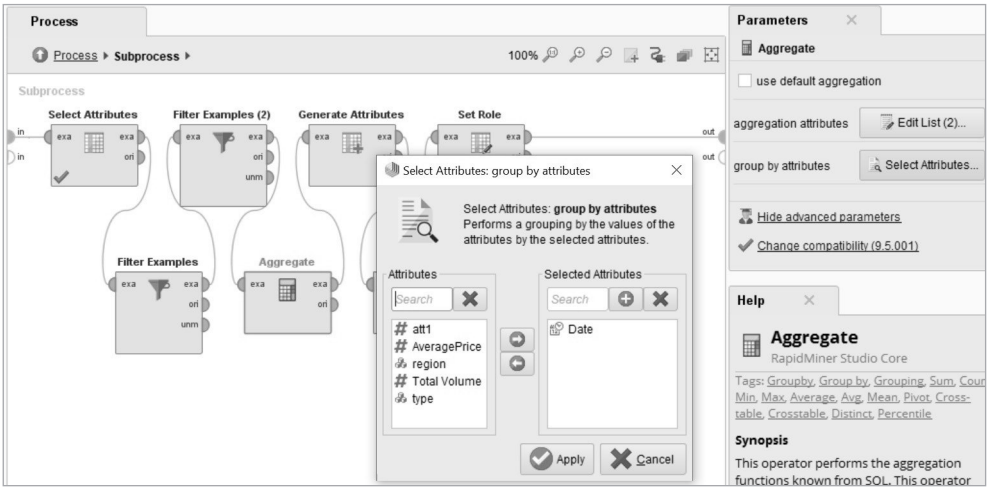


Figura D.25 – Aggregate, per selezionare l'attributo (Date) secondo cui aggregare i dati.

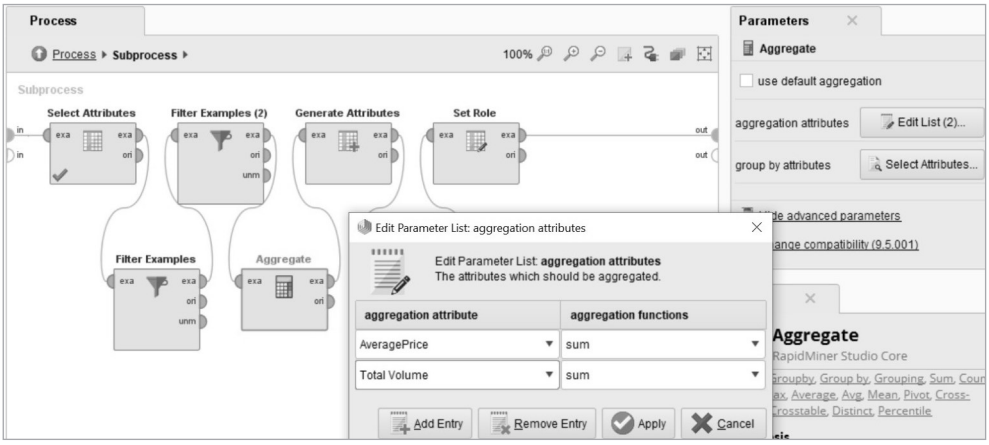


Figura D.26 – Aggregate, operazioni di somma sugli attributi da aggregare.

Si potrebbe scegliere subito Aggregation Function con valore average, ma è meglio mostrare come si introducono funzioni più complesse. Inoltre, Aggregate seleziona gli attributi, dunque il suo utilizzo rende ridondante l'uso del primo SelectAttribute, introdotto a titolo illustrativo.

Per ottenere la media di AveragePrice bisogna dividere per il numero di distributori pari a 54, cioè bisogna eseguire le divisioni sui valori precedenti; lo stesso procedimento va eseguito per TotalVolume. Le operazioni da svolgere riguardano GenerateAttributes, poi Parameters; generare la lista degli attributi da sostituire con le specifiche operazioni di divisione, modificando il campo Function description con Function ex-

pression. Si noti che viene assegnato un nuovo nome all'attributo su cui si opera, come mostrato nelle Figure D.27 e D.28.

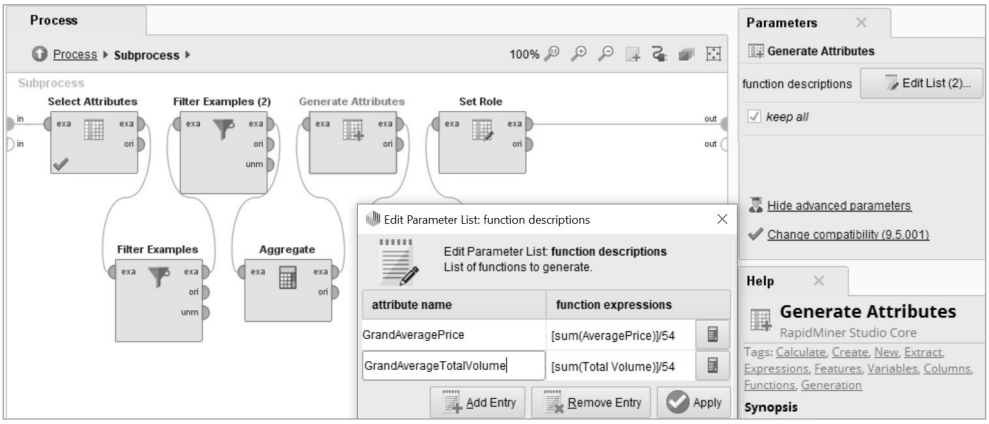


Figura D.27 – GenerateAttribute, trasformazioni aritmetiche e statistiche sui dati.

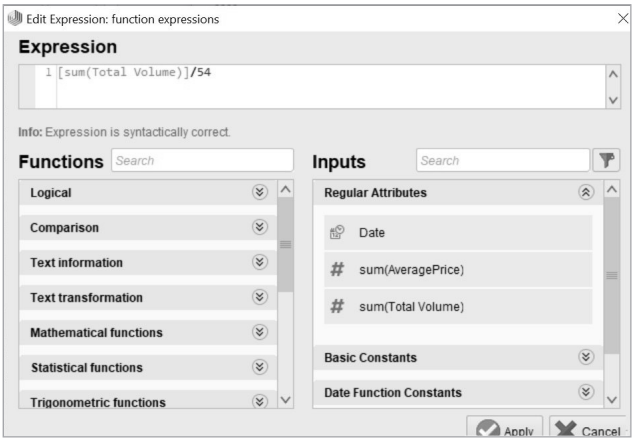


Figura D.28 – Inserimento delle espressioni algebriche.

Trattandosi di una serie temporale, bisogna accertarsi che tutte le occorrenze siano in ordine cronologico, per individuare in modo corretto quello che, come spiegato più avanti, viene selezionato come training e test set. Questa operazione è superflua per questo specifico caso, dal momento che le osservazioni risultano già nel giusto ordine; in generale è un'operazione che previene errori dovuti alla fase di rilevamento dei dati. Le operazioni da svolgere riguardano Sort, poi Parameters; scegliere la variabile secondo cui riordinare le occorrenze con il campo Attribute name, scegliere il tipo di ordinamento con il campo Sorting direction, come mostrato in Figura D.29.

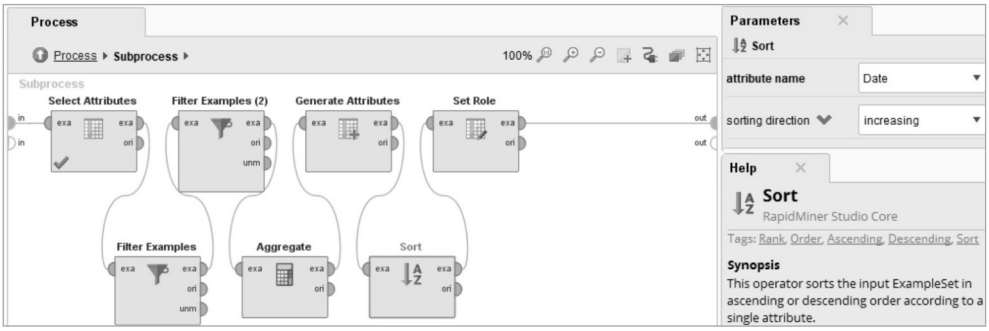


Figura D.29 – Sort, riordinamento temporale.

Infine, assegnare il ruolo di variabile target all'attributo AveragePrice; questo viene fatto assegnando all'attributo selezionato l'etichetta label. Le operazioni da svolgere riguardano SetRole, poi Parameters; scegliere l'attributo AveragePrice con il campo Attribute name; attribuire a questo il ruolo di variabile target con il campo Target role con valore label, come mostrato in Figura D.30.

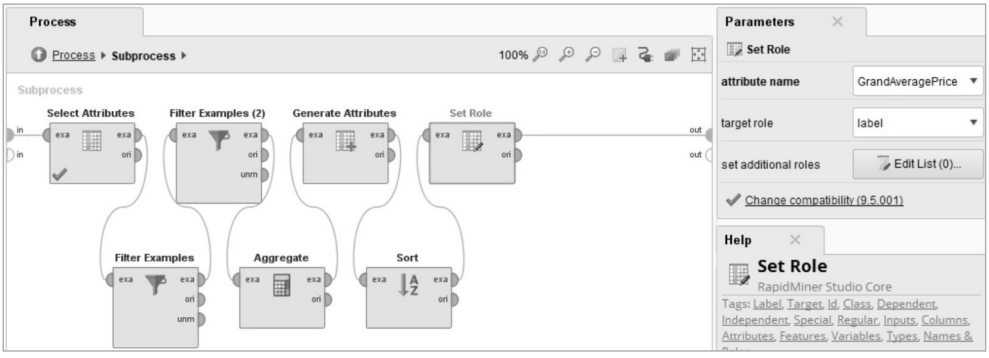


Figura D.30 – SetRole, assegnazione della variabile target.

Finalmente si può passare a addestrare il modello di previsione e successivamente applicarlo ai dati per predire la variabile target AveragePrice.

Training, Validation e Test del Modello

Per fare previsioni su AveragePrice si può adottare il modello di neural network MultiLayer Perceptron (MLP) con DeepLearning, che offre la possibilità di scegliere un maggiore numero di parametri rispetto a Neural Net.

Per eseguire un controllo e selezionare la combinazione più efficiente dei vari parametri di tuning dell'intero modello, si può usare l'operatore OptimizeParameters(Grid) che, come un operatore di sottoprocesso, incapsula al suo interno tutti gli operatori

utili alle varie fasi. Bisogna inserire OptimizeParameters(Grid) nel processo principale, in serie all'operatore Subprocess, come mostrato in Figura D.31.

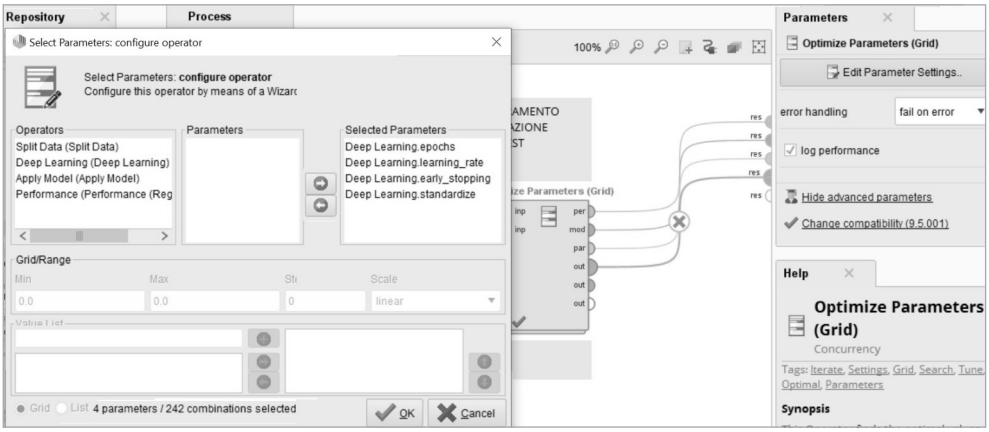


Figura D.31 – OptimizeParameters(Grid), selezione dei parametri da confrontare.

Tramite OptimizeParameters(Grid) e Parameters si può scegliere la variabile secondo cui riordinare le occorrenze con il campo Attribute name, e scegliere il tipo di ordinamento modificando il campo Sorting direction, come mostrato in Figura D.32.

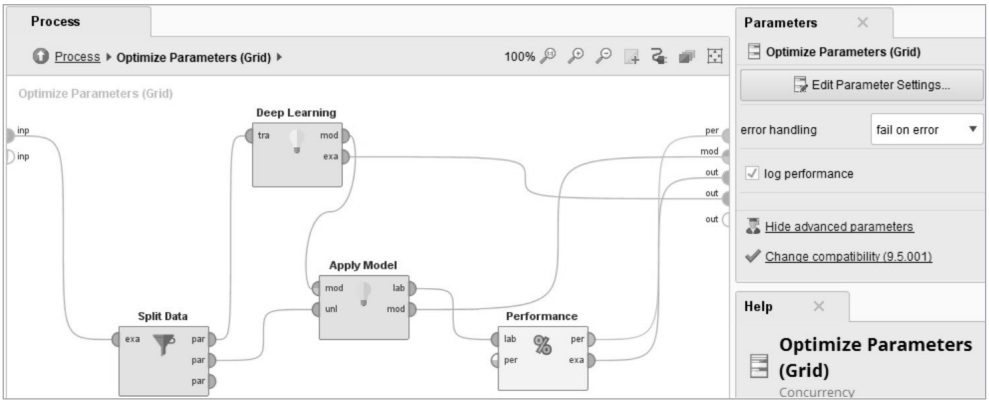


Figura D.32 – Sottoprocesso interno a OptimizeParameters(Grid).

Si procede con il disegnare i processi relativi all'analisi e alla configurazione di operatori. Con il dataset caricato vengono generati il training set con il 95% e il test set con il 5%; le percentuali sono scelte perché, trattandosi di una serie temporale, abbiamo bisogno di molti dati in fase di addestramento e la previsione è tanto più accurata quanto più è a breve termine. Serve attenzione nel rispettare l'ordine temporale nella

separazione tra training e test, ovvero nel considerare una separazione lineare dei dataset, in quanto una scelta casuale delle occorrenze romperebbe la coerenza temporale. Le operazioni da svolgere riguardano SplitData, poi Parameters; scegliere il tipo di campionamento con il campo Sampling type con valore linear, creare le due partizioni con il campo Partition, come in Figura D.33.

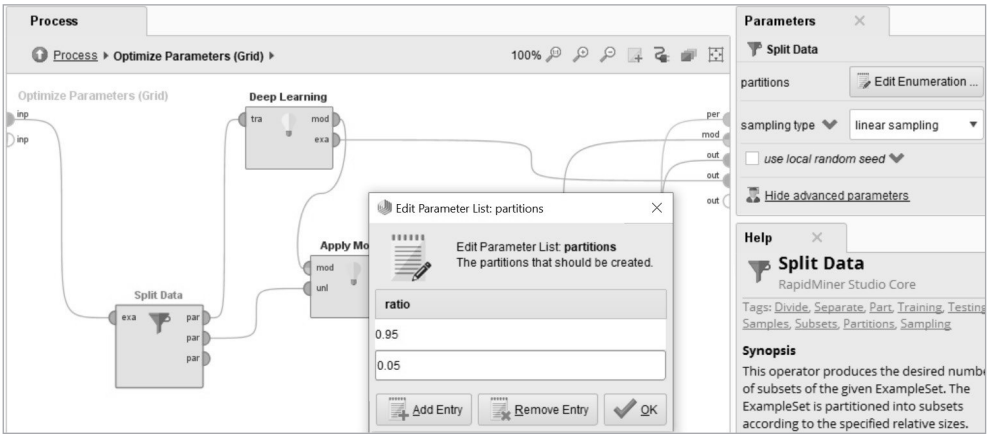


Figura D.33 – SplitData, generazione dei due dataset di training e di test.

Viene inserito DeepLearning collegandolo all'out di SplitData associata alla porzione di dati corretta; l'ordine delle porte in SplitData è lo stesso dei tab che contengono le porzioni di dati considerate. In questo modo il 95% dei dati andrà a addestrare DeepLearning, mentre il restante 5% sarà riservato al test. Come detto, DeepLearning implementa una rete multistrato i cui parametri vanno opportunamente modificati per ottimizzare la previsione. Come esercizio, si possono verificare le diverse combinazioni dei parametri. Le operazioni da svolgere riguardano DeepLearning, poi Parameters; scegliere il numero di strati e di nodi con il campo Hidden layer sizes, scegliere tre strati con un numero di nodi rispettivamente pari a 60, 120, 60, modificare il campo L1 con 0,3 e L2 con 0,2, come in Figura D.34.
Una volta addestrato DeepLearning, si può eseguirlo su dati di test non ancora presentati al modello. Bisogna collegare ApplyModel in serie a SplitData attraverso la porta unl e in parallelo con DeepLearning attraverso la porta mod. Questo permette di passare a ApplyModel la parte restante dei dati di test da SplitData e i parametri addestrati da DeepLearning. Da notare che l'out mod di ApplyModel viene collegato all'out interno di OptimizeParameters(Grid). Una volta collegato l'out di quest'ultimo operatore all'out del processo principale, l'out mod rappresenta l'uscita da cui ottenere la previsione cercata. Per ApplyModel e Parameters si lasciano i valori di default.

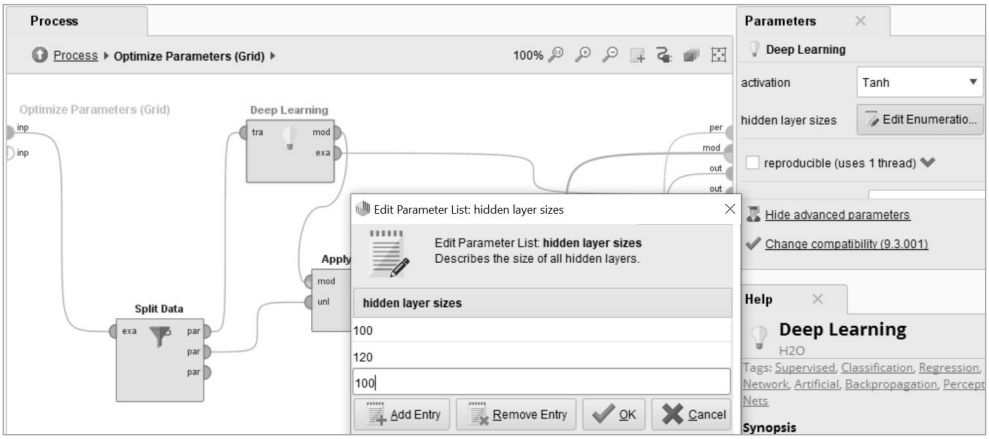


Figura D.34 – DeepLearning, il modello di machine learning scelto per la previsione.

Infine, viene collegato in serie a ApplyModel, attraverso le porte lab, l'operatore Performance, che misura l'efficienza della previsione, sulla base di una serie di indicatori selezionabili da Parameters. Viene scelto Performance(Regression), avendo a che fare con serie storiche. L'out per e l'out exa di Performance vengono entrambi collegati all'out interno di OptimizeParameters(Grid), e da qui alle uscite del processo completo; verranno quindi visualizzati in output tra i risultati dell'analisi.
In merito a Performance e Parameters, scegliere gli indicatori di accuratezza da mostrare in output con i campi root mean square error, absolute error, relative error, come mostrato in Figura D.35.

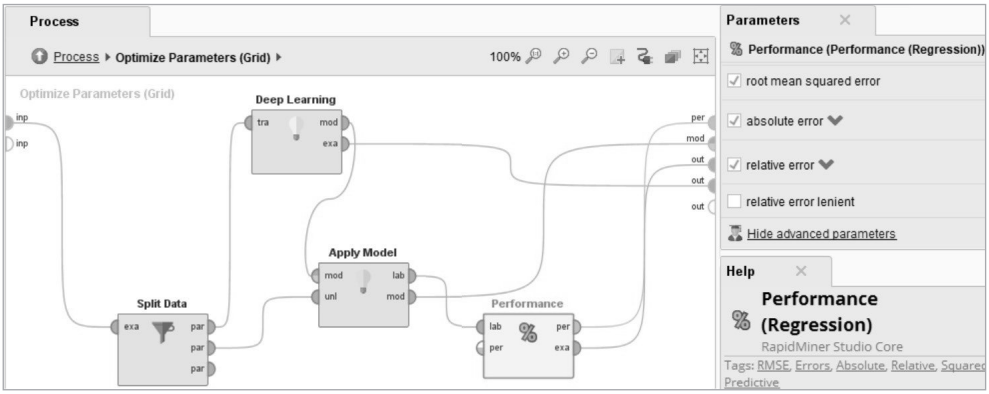
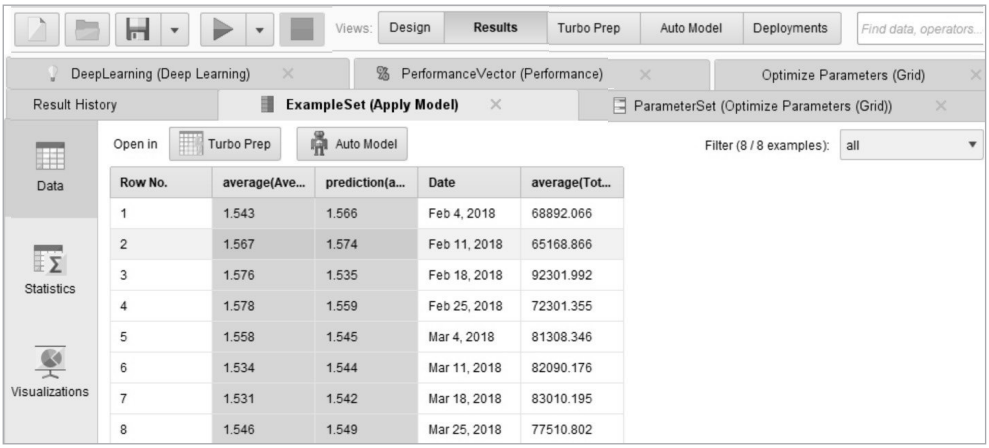


Figura D.35 – Performance, per ottenere gli indicatori di accuratezza della previsione.

Bisogna tornare al Design del processo principale, avviare l'esecuzione con run, l'area Results mostra i risultati del test, del training e di performance del modello.

Il test set contiene le occorrenze (in settimane) con i valori dell'attributo Date compresi tra il 4 febbraio e il 25 marzo 2018. Come si vede dalle figure seguenti, i risultati sono incoraggianti: l'errore relativo risulta essere l'1,03% e lo scarto quadratico medio pari a 0,020.



Row No.	average(Ave...	prediction(a...	Date	average(Tot...
1	1.543	1.566	Feb 4, 2018	68892.066
2	1.567	1.574	Feb 11, 2018	65168.866
3	1.576	1.535	Feb 18, 2018	92301.992
4	1.578	1.559	Feb 25, 2018	72301.355
5	1.558	1.545	Mar 4, 2018	81308.346
6	1.534	1.544	Mar 11, 2018	82090.176
7	1.531	1.542	Mar 18, 2018	83010.195
8	1.546	1.549	Mar 25, 2018	77510.802

Figura D.36 – Risultati con il test set: la prima colonna da destra indica i valori realmente rilevati, la seconda i valori previsti dal modello.

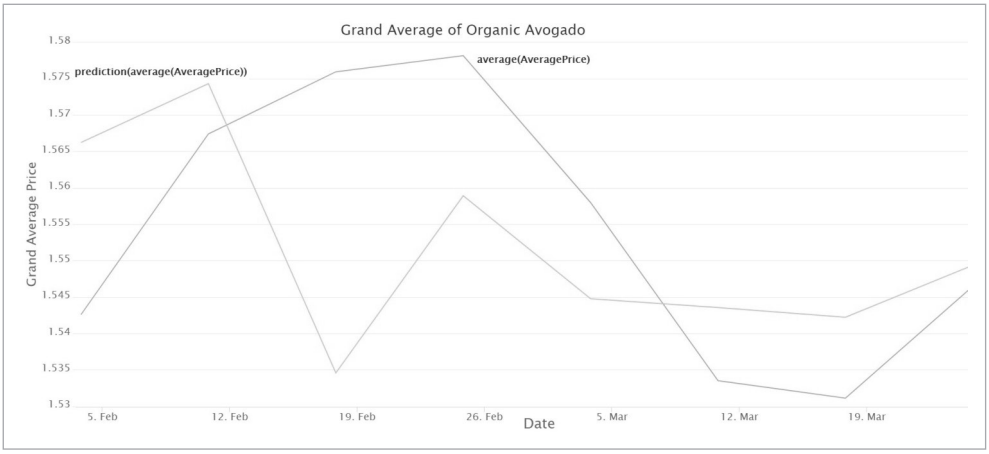


Figura D.37 – Valori effettivi e valori previsti dal 4 febbraio al 25 marzo 2018.

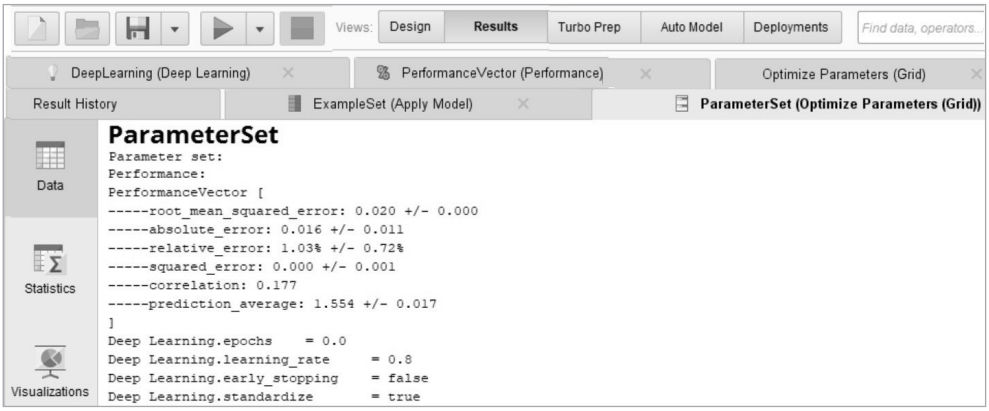


Figura D.38 – Indicatori di efficienza ottenuti dall'operatore Performance.

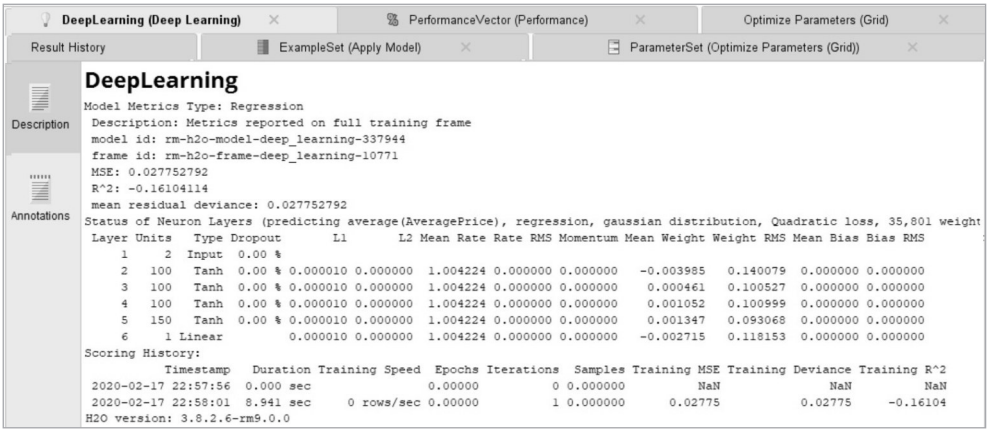


Figura D.39 – Rapporto delle iterazioni e dei parametri utilizzati da DeepLearning.

Riferimenti bibliografici

A. Chisholm, *Exploring data with RapidMiner*, Packt Publishing, Birmingham (UK) 2013.

M. Hofmann, R. Klinkenberg, *RapidMiner*: *Data Mining use cases and business analytics applications*, Taylor & Francis Ltd, Abingdon (UK) 2019, seconda edizione.

Note

1 Rodolfo Baggio, https://it.linkedin.com/in/rbaggio
2 Filippo Carone Fabiani, https://it.linkedin.com/in/filippo-carone-fabiani-8823877
3 RapidMiner, https://rapidminer.com/
4 RapidMiner pricing, https://rapidminer.com/pricing/
5 RapidMiner Studio: Start your free 30-day trial, https://rapidminer.com/get-started/
6 RapidMiner Educational License Program, https://rapidminer.com/educational-program/
7 Feedback on RapidMiner vs. Python comparison, https://community.rapidminer.com/discussion/55155/feedback-on-rapidminer-vs-python-comparison
8 *RapidMiner vs R? How to use Python and R together with RapidMiner,* https://rapidminer.com/blog/rapidminer-vs-r-tips-tricks-how-to-use-python-r-rapidminer/
9 https://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf e https://docs.rapidminer.com/downloads/DataMiningForTheMasses.pdf
10 RapidMiner Documentation, https://docs.rapidminer.com/
11 Spambase Data Set, https://archive.ics.uci.edu/ml/datasets/spambase
12 arules: Mining Association Rules and Frequent Itemsets, https://cran.r-project.org/web/packages/arules/
13 Text Processing, https://marketplace.rapidminer.com/UpdateServer/faces/product_details.xhtml?productId=rmx_text
14 Hotel Reviews, https://www.kaggle.com/datafiniti/hotel-reviews#Datafiniti_Hotel_Reviews.csv
15 Avocado Prices, https://www.kaggle.com/neuromusic/avocado-prices

Dati • Machine learning • Neural network • Deep learning

Che cos'è l'intelligenza artificiale? Come creare un algoritmo per risolvere problemi computazionali complessi? Quali sono i vantaggi e gli svantaggi? Come organizzare i dati? Come preparare l'input e interpretare l'output? Come scegliere le librerie e gli strumenti di programmazione?
Questo libro intende rispondere a queste e altre domande con un approccio pragmatico orientato al "ragionare per trovare soluzioni". Rivolto al programmatore che vuole avviare lo sviluppo degli algoritmi, è utile anche a chi desidera capire come funzionano certe soluzioni o immaginare nuovi utilizzi.
Il volume è ricco di esempi, consigli, codice in linguaggio Python e link, selezionati con cura per cominciare subito a sperimentare gli approcci principali e conoscere le problematiche esistenti.

Roberto Marmo è professore a contratto di informatica presso la facoltà di Ingegneria dell'Università degli Studi di Pavia, consulente e formatore sull'intelligenza artificiale per cercare e analizzare informazioni estratte da internet, social media e altre fonti. È autore di altri libri, fra cui *Social Media Mining* e *Matematica Rock*, entrambi pubblicati da Hoepli.
www.robertomarmo.net

Tra gli argomenti trattati:

- analisi economico-finanziaria delle aziende;
- progettazione dell'algoritmo;
- algoritmo evolutivo, logica fuzzy, sistema esperto;
- vari modelli di machine learning;
- neural network e deep learning;
- analisi di smartphone con IoT e sales forecast.

Risorse online
Su www.algoritmiia.it e www.hoeplieditore.it/9171-3 sono disponibili numerosi materiali gratuiti a integrazione del testo.

www.hoeplieditore.it
Ulrico Hoepli Editore S.p.A.
via Hoepli, 5 - 20121 Milano
e-mail hoepli@hoepli.it

€ 29,90

e book disponibile

ISBN 978-88-203-9171-3