



Roll No.: 19BCE245

Course Code and Name: **2CS501 MACHINE LEARNING**

## Practical: 2

```
[1] 1 pip install camelot-py[cv] tabula-py[cv]
```

```
Collecting camelot-py[cv]
  Downloading camelot_py-0.10.1-py3-none-any.whl (40 kB)
    [REDACTED] 40 kB 11 kB/s
Collecting tabula-py[cv]
  Downloading tabula_py-2.3.0-py3-none-any.whl (12.0 MB)
    [REDACTED] 12.0 MB 96 kB/s
Requirement already satisfied: tabulate>=0.8.9 in /usr/local/lib/python3.7/dist-packages (from camelot-py[cv]) (0.8.9)
Requirement already satisfied: chardet>=3.0.4 in /usr/local/lib/python3.7/dist-packages (from camelot-py[cv]) (3.0.4)
Requirement already satisfied: pandas>=0.23.4 in /usr/local/lib/python3.7/dist-packages (from camelot-py[cv]) (1.1.5)
Collecting PyPDF2>=1.26.0
  Downloading PyPDF2-1.26.0.tar.gz (77 kB)
    [REDACTED] 77 kB 5.7 MB/s
Requirement already satisfied: click>=6.7 in /usr/local/lib/python3.7/dist-packages (from camelot-py[cv]) (7.1.2)
Requirement already satisfied: openpyxl>=2.5.8 in /usr/local/lib/python3.7/dist-packages (from camelot-py[cv]) (2.5.9)
Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.7/dist-packages (from camelot-py[cv]) (1.19.5)
Collecting pdfminer.six>=20200726
  Downloading pdfminer.six-20201018-py3-none-any.whl (5.6 MB)
    [REDACTED] 5.6 MB 64.4 MB/s
Requirement already satisfied: opencv-python>=3.4.2.17 in /usr/local/lib/python3.7/dist-packages (from camelot-py[cv]) (4.1.2.20)
Collecting ghostscript>=0.7
  Downloading ghostscript-0.7-py2.py3-none-any.whl (25 kB)
Collecting pdftopng>=0.2.3
  Downloading pdftopng-0.2.3-cp37m-manylinux2010_x86_64.whl (11.7 MB)
    [REDACTED] 11.7 MB 22.3 MB/s
Requirement already satisfied: setuptools>=38.6.0 in /usr/local/lib/python3.7/dist-packages (from ghostscript>=0.7->camelot-py[cv]) (50.3.2)
Requirement already satisfied: et-xmlfile in /usr/local/lib/python3.7/dist-packages (from openpyxl>=2.5.8->camelot-py[cv]) (1.0.4)
Requirement already satisfied: jdcal in /usr/local/lib/python3.7/dist-packages (from openpyxl>=2.5.8->camelot-py[cv]) (1.4.6)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.23.4->camelot-py[cv]) (2.8.1)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.23.4->camelot-py[cv]) (2019.1)
Collecting cryptography
  Downloading cryptography-3.4.8-cp36-abi3-manylinux2_24_x86_64.whl (3.0 MB)
    [REDACTED] 3.0 MB 56.3 MB/s
Requirement already satisfied: sortedcontainers in /usr/local/lib/python3.7/dist-packages (from pdfminer.six>=20200726->camelot-py[cv]) (4.4.4)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from python-dateutil>=2.7.3->pandas>=0.23.4->camelot-py[cv]) (1.14.0)
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.7/dist-packages (from cryptography->pdfminer.six>=20200726->camelot-py[cv]) (1.14.0)
Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages (from cffi>=1.12->cryptography->pdfminer.six>=20200726->camelot-py[cv]) (2.0)
WARNING: tabula-py 2.3.0 does not provide the extra 'cv'
Collecting distro
  Downloading distro-1.6.0-py2.py3-none-any.whl (19 kB)
Building wheels for collected packages: PyPDF2
  Building wheel for PyPDF2 (setup.py) ... done
  Created wheel for PyPDF2: filename=PyPDF2-1.26.0-py3-none-any.whl size=61101 sha256=82a007899fb5208ea593b9c90e4b11be9692f
  Stored in directory: /root/.cache/pip/wheels/80/1a/24/648467ade3a77ed20f35cfd2badd32134e96dd25ca811e64b3
Successfully built PyPDF2
Installing collected packages: cryptography, PyPDF2, pdfminer.six, distro, tabula-py, pdftopng, ghostscript, camelot-py
Successfully installed PyPDF2-1.26.0 camelot-py-0.10.1 cryptography-3.4.8 distro-1.6.0 ghostscript-0.7 pdfminer.six-20201018
```

```
[2] 1 !sudo apt install ghostscript
```

```

Reading package lists... Done
Building dependency tree
Reading state information... Done
The following package was automatically installed and is no longer required:
  libnvidia-common-460
Use 'sudo apt autoremove' to remove it.
The following additional packages will be installed:
  fonts-droid-fallback fonts-noto-mono gsfonts libcupsfilters1 libcupsimage2
  libgs9 libgs9-common libijs-0.35 libjbig2dec0 poppler-data
Suggested packages:
  fonts-noto ghostscript-x poppler-utils fonts-japanese-mincho
  | fonts-ipafont-mincho fonts-japanese-gothic | fonts-ipafont-gothic
  fonts-arphic-ukai fonts-arphic-uming fonts-nanum
The following NEW packages will be installed:
  fonts-droid-fallback fonts-noto-mono ghostscript gsfonts libcupsfilters1
  libcupsimage2 libgs9 libgs9-common libijs-0.35 libjbig2dec0 poppler-data
0 upgraded, 11 newly installed, 0 to remove and 40 not upgraded.
Need to get 14.1 MB of archives.
After this operation, 49.9 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-droid-fallback all 1:6.0.1r16-1.1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu bionic/main amd64 poppler-data all 0.4.8-2 [1,479 kB]
Get:3 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-noto-mono all 20171026-2 [75.5 kB]
Get:4 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcupsimage2 amd64 2.2.7-1ubuntu2.8 [18.6 kB]
Get:5 http://archive.ubuntu.com/ubuntu bionic/main amd64 libijs-0.35 amd64 0.35-13 [15.5 kB]
Get:6 http://archive.ubuntu.com/ubuntu bionic/main amd64 libjbig2dec0 amd64 0.13-6 [55.9 kB]
Get:7 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libgs9-common all 9.26-dfsg+0-0ubuntu0.18.04.14 [5,092 kB]
Get:8 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libgs9 amd64 9.26-dfsg+0-0ubuntu0.18.04.14 [2,265 kB]
Get:9 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 ghostscript amd64 9.26-dfsg+0-0ubuntu0.18.04.14 [51.3 kB]
Get:10 http://archive.ubuntu.com/ubuntu bionic/main amd64 gsfonts all 1:8.11+urwcyr1.0.7-pre44-4.4 [3,120 kB]
Get:11 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcupsfilters1 amd64 1.20.2-0ubuntu3.1 [108 kB]
Fetched 14.1 MB in 1s (22.5 MB/s)
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based frontend cannot be used. at /usr/share/perl5/Debconf)
debconf: falling back to frontend: Readline
debconf: unable to initialize frontend: Readline
debconf: (This frontend requires a controlling tty.)
debconf: falling back to frontend: Teletype
dpkg-preconfigure: unable to re-open stdin:
Selecting previously unselected package fonts-droid-fallback.
(Reading database ... 148486 files and directories currently installed.)

```

```
Preparing to unpack .../00-fonts-droid-fallback 1:6.0.1r16-1.1_all.deb ...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../01-poppler-data_0.4.8-2_all.deb ...
Unpacking poppler-data (0.4.8-2) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../02-fonts-noto-mono_20171026-2_all.deb ...
Unpacking fonts-noto-mono (20171026-2) ...
Selecting previously unselected package libcupsimage2:amd64.
Preparing to unpack .../03-libcupsimage2_2.2.7-1ubuntu2.8_amd64.deb ...
Unpacking libcupsimage2:amd64 (2.2.7-1ubuntu2.8) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../04-libijs-0.35_0.35-13_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-13) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../05-libjbig2dec0_0.13-6_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.13-6) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../06-libgs9-common_9.26-dfsg+0-0ubuntu0.18.04.14_all.deb ...
Unpacking libgs9-common (9.26-dfsg+0-0ubuntu0.18.04.14) ...
```

```
✓ 1s [3] 1 import camelot
      2 tables = camelot.read_pdf('apple_report_q1.pdf')
```

```
✓ 0s [4] 1 tables
```

```
<TableList n=1>
```

```
✓ 0s [5] 1 tables[0]
```

```
<Table shape=(28, 3)>
```

```
✓ 0s [6] 1 tables.export('web_report.html',f='html')
```

```
✓ 0s [7] 1 tables[0].to_html('web_report_1.html')
      2 tables[0].to_json('json_report.json')
      3 tables[0].to_excel('excel_report.xlsx')
      4 tables[0].to_sqlite('sql_report.sqlite')
```

```
✓ 2s [8] 1 import tabula
      2 tables_2 = tabula.read_pdf("apple_report_q1.pdf", pages='all')
```

```
✓ 0s 1 tables_2
```

```
[
0      Apple Inc.
1  CONDENSED CONSOLIDATED STATEMENTS OF OPERATION...
2  (In millions, except number of shares which ar...
3      Three months ended
4      December 26, 2020  December 28, 2019
5      Net sales
6      Products  95,678  79,104
7      Services  15,761  12,715
8      Total net sales  1,11,439  91,819
9      Operating Income  33,534  25,569
10     Other income/ expense, net  45 349
11     Income before provision for income taxes  33,57...
12     Provision for income taxes  4,824  3,682
13     Net income  28,755  22,236
14     Earnings per share
15     Basic  1.70  1.26
16     Diluted  1.68  1.25
17     Shares used in computing earnings per share
18     Basic  1,69,35,119  1,76,60,160
19     Diluted  1,71,13,688  1,78,18,417
20     Net sales by reportable segment :
21     Americas  46,310  41,367
22     Europe  27,306  23,273
23     Greater China  21,313  13,578
24     Japan  8,225  7,378
25     Rest of Asia pacific  8,225  7,378
      Total net sales  1,11,439  91,819]
```