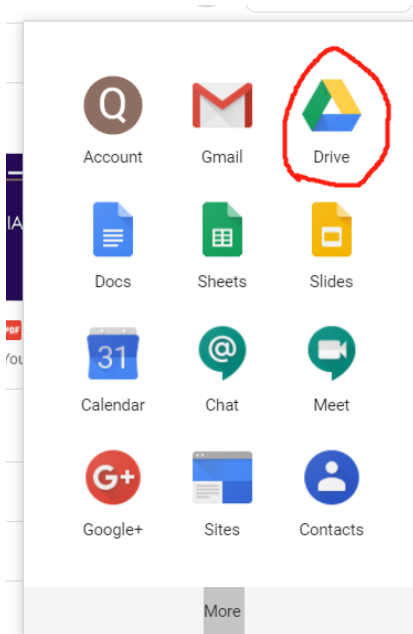
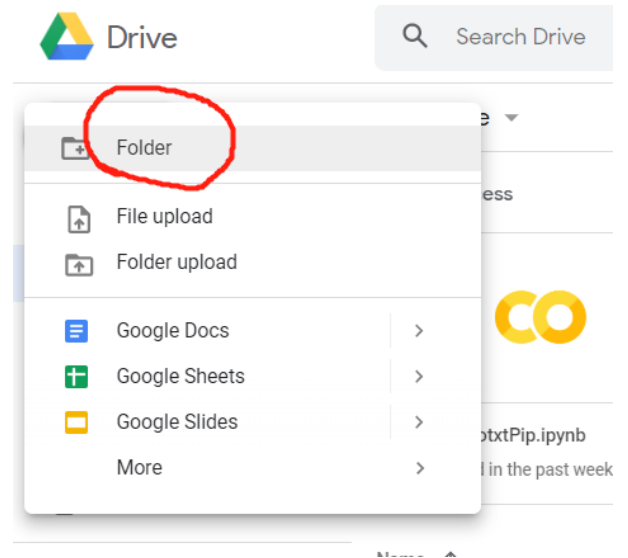


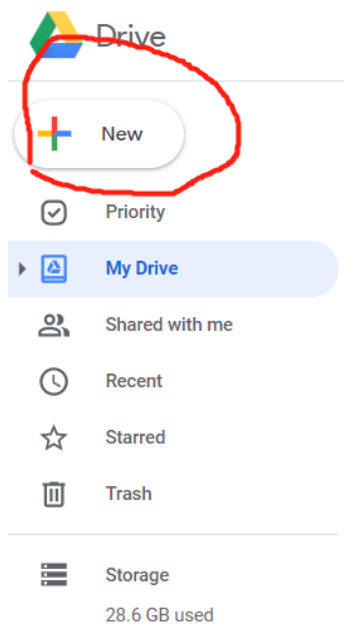
Step1:



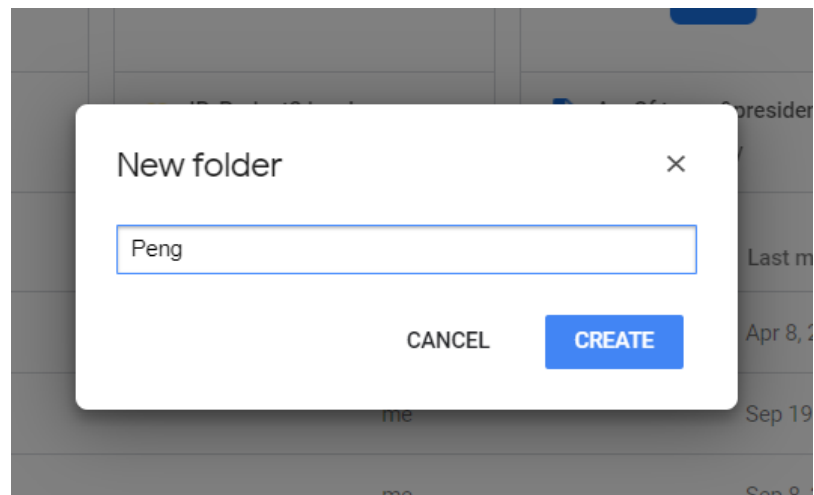
Step3:



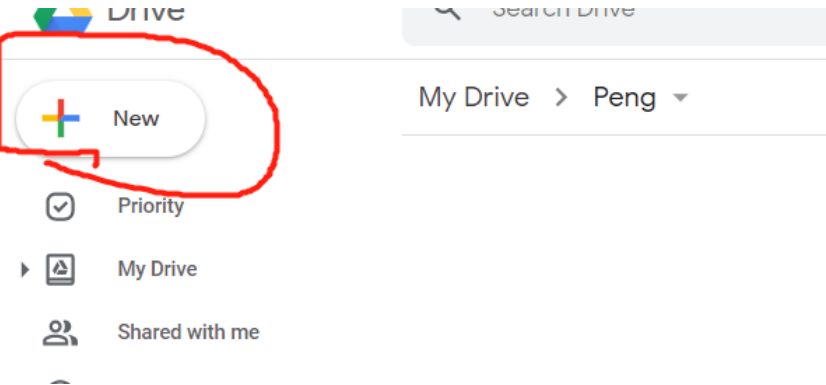
Step2:



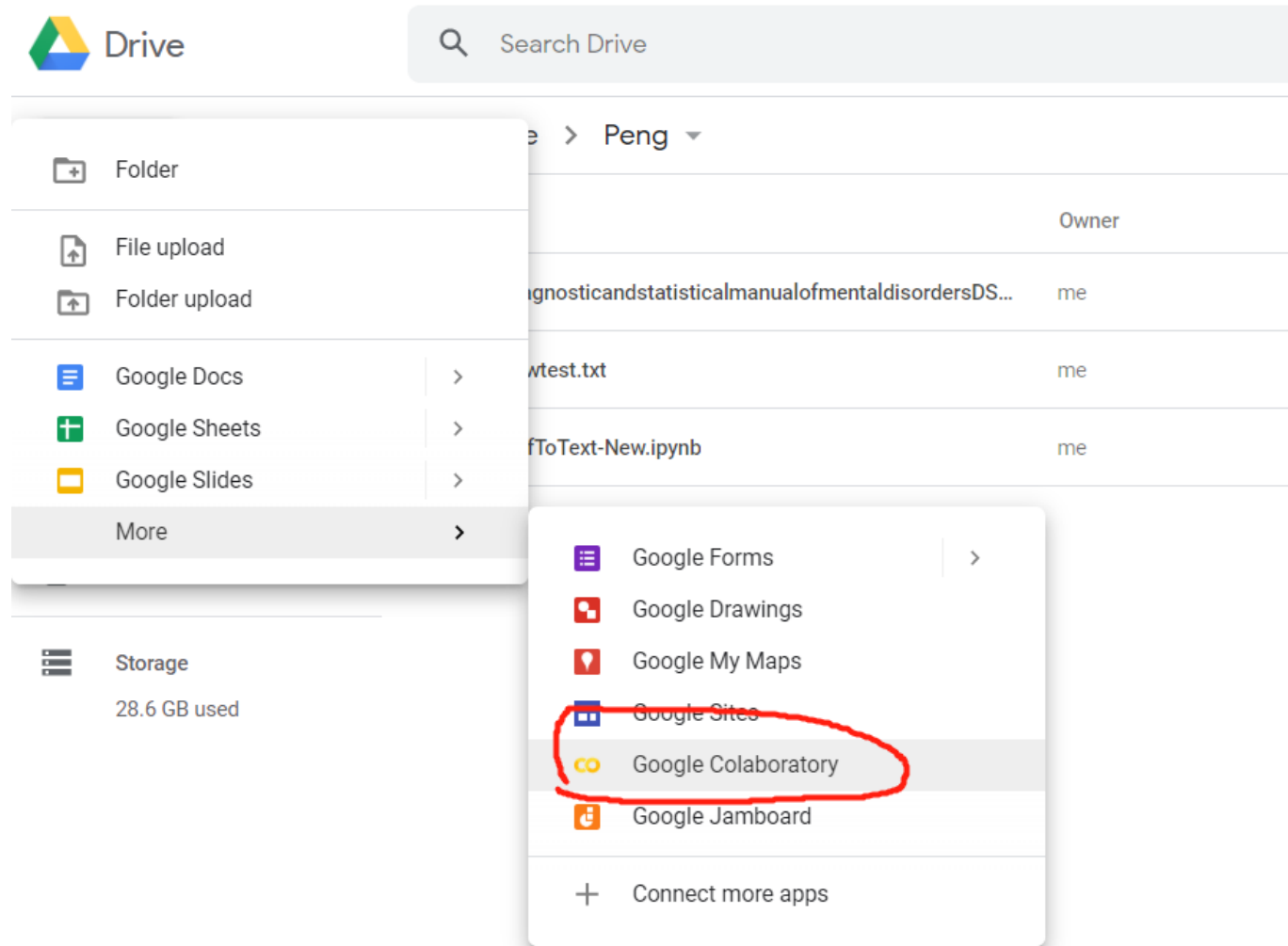
Step4:



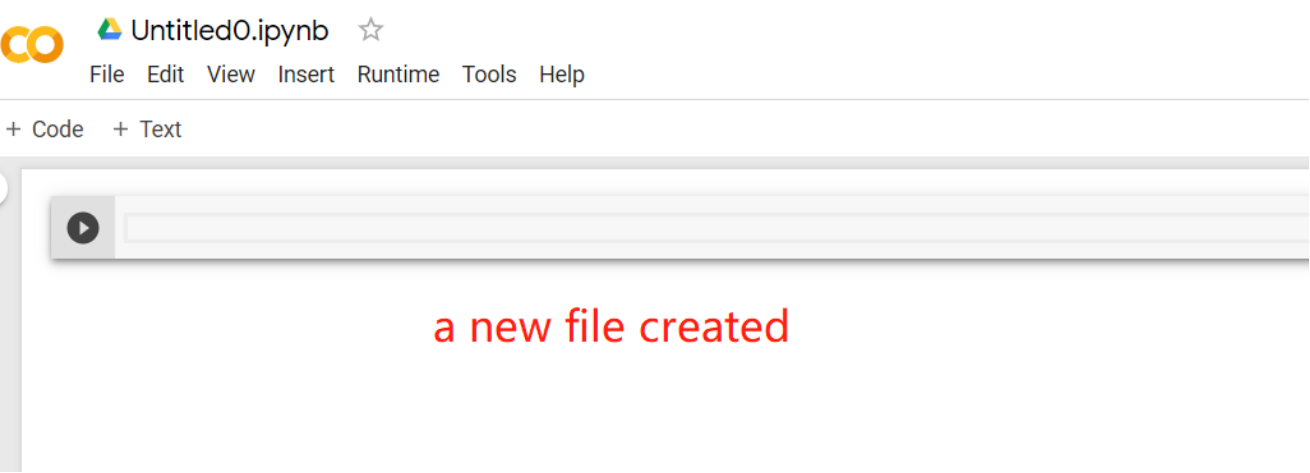
Step 5



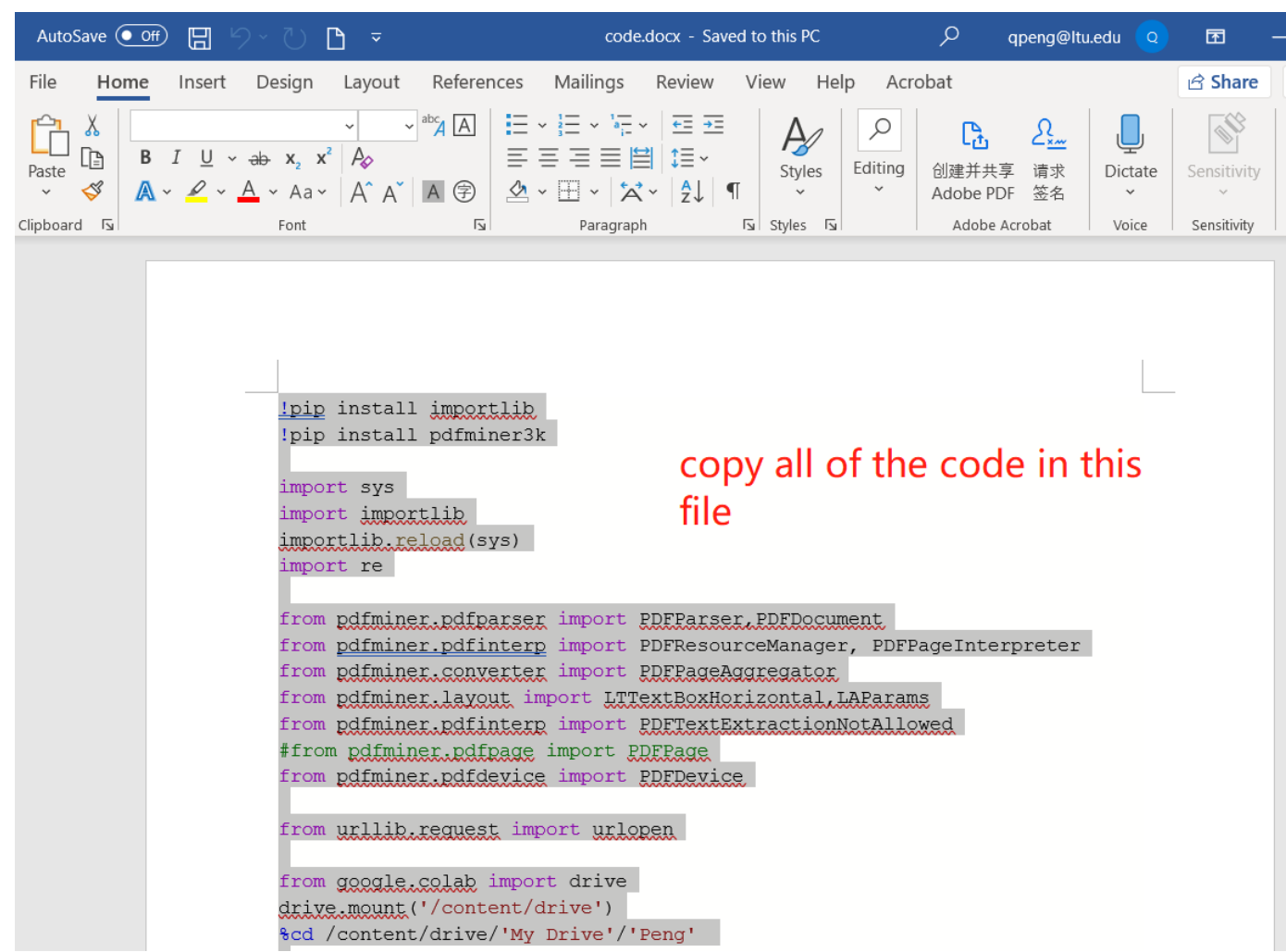
Step 6




Step 7




Step 8




Step 9

 **Untitled0.ipynb** ☆
File Edit View Insert Runtime Tools Help [All changes saved](#)


Code + Text

 |

paste all the code in here

 **Untitled0.ipynb** ☆
File Edit View Insert Runtime Tools Help [All changes saved](#)

+ Code + Text



```
!pip install importlib
!pip install pdfminer3k

import sys
import importlib
importlib.reload(sys)
import re

from pdfminer.pdfparser import PDFParser,PDFDocument
from pdfminer.pdfinterp import PDFResourceManager, PDFPageInterpreter
from pdfminer.converter import PDFPageAggregator
from pdfminer.layout import LTTextBoxHorizontal,LAParams
from pdfminer.pdfinterp import PDFTextExtractionNotAllowed
#from pdfminer.pdfpage import PDFPage
from pdfminer.pdfdevice import PDFDevice

from urllib.request import urlopen

from google.colab import drive
drive.mount('/content/drive')
%cd /content/drive/'My Drive'/'Peng'

pdf_path = 'DiagnosticandstatisticalmanualofmentaldisordersDSM-5.pdf'
newname1 = 'newtest.txt'

def extract_layout_by_page(pdf_path):

    laparams = LAParams() # pdfminer.layout.LAParams(line_overlap=0.5, char_margin=2.0, line_margin=0.5, word_mar

    fp = open(pdf_path, 'rb') #
    parser = PDFParser(fp) # Create a PDF parser object associated with the file object.
    document = PDFDocument(parser) # Create a PDF document object that stores the document structure.

    #Connect analyzer to document object
    parser.set_document(document)
```

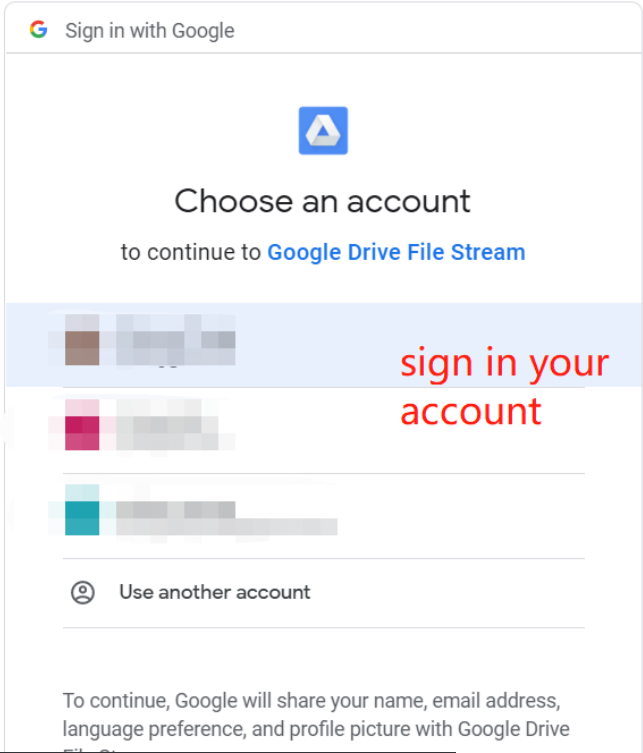
+ Code+ Text

RAMDisk

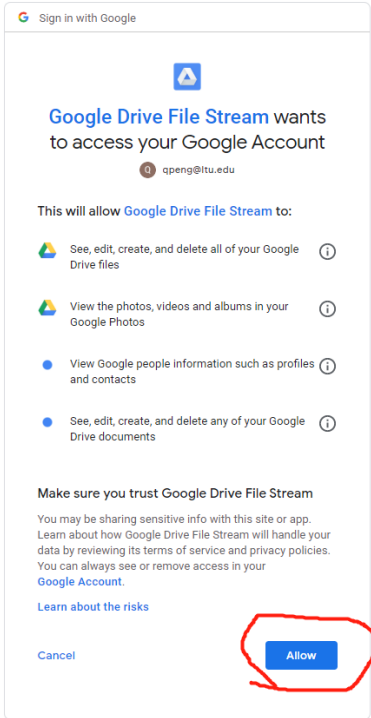
Editing

```
Created wheel for importlib: filename=importlib-1.0.4-cp36-none-any.whl size=5858 sha256=6c405d5a878bdb10189f058dd34919a5717b7bc557f26672
Stored in directory: /root/.cache/pip/wheels/26/eb/96/a8a05637794153edb35b7f7f97de528350367d42d8829c187
...
Successfully built importlib
Installing collected packages: importlib
Successfully installed importlib-1.0.4
Collecting pdfminer3k
  Downloading https://files.pythonhosted.org/packages/8c/87/cee0aa24f95c287020df7e3936cb51d32b34b05b430759bac15f89ea5ac2/pdfminer3k-1.3.1.t
  [redacted] | 4.1MB 3.5MB/s
Requirement already satisfied: pytest>=2.0 in /usr/local/lib/python3.6/dist-packages (from pdfminer3k) (3.6.4)
Collecting ply>=3.4
  Downloading https://files.pythonhosted.org/packages/a3/58/35da89ee790598a0700ea49b2a66594140f44dec458c07e8e3d4979137fc/ply-3.11-py2.py3-n
  [redacted] | 51kB 7.4MB/s
Requirement already satisfied: atomicwrites>=1.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (1.3.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (42.0.1)
Requirement already satisfied: py>=1.5.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (1.8.0)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (19.3.0)
Requirement already satisfied: more-itertools>=4.0.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (7.2.0)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (1.12.0)
Requirement already satisfied: pluggy<0.8,>=0.5 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (0.7.1)
Building wheels for collected packages: pdfminer3k
  Building wheel for pdfminer3k (setup.py) ... done
  Created wheel for pdfminer3k: filename=pdfminer3k-1.3.1-cp36-none-any.whl size=105356 sha256=a23e1d10fdda019d176374c394fbe83596c2f07a98a6
  Stored in directory: /root/.cache/pip/wheels/ca/4f/a7/cb601b4fb257d2321ac668b7c6e269176780bd0283eda855d2
Successfully built pdfminer3k
Installing collected packages: ply, pdfminer3k
Successfully installed pdfminer3k-1.3.1 ply-3.11
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googlele
Enter your authorization code:
click this link
```

Step 10



Step 11



Step 12



Sign in

Please copy this code, switch to your application and paste it there

4/twEn601 QRVQ8r1lP4DyhCMLBAmdWtp9Zf2MZXE4bdXAaoDP
2qWnqNk



click this to cope
the code

Step 13

Step 14

paste the code here and enter in you
keyboard

Step 15

PdfToText-New.ipynb

Comment Share

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

RAM Disk Editing

```
Downloading https://files.pythonhosted.org/packages/a3/58/35da89ee790598a0700ea49b2a66594140f44dec458c07e8e3d4979137fc/ply-3.11-py2.py3-nu
51kB 7.4MB/s
Requirement already satisfied: atomicwrites>=1.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (1.3.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (42.0.1)
Requirement already satisfied: py>=1.5.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (1.8.0)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (19.3.0)
Requirement already satisfied: more-itertools>=4.0.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (7.2.0)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (1.12.0)
Requirement already satisfied: pluggy<0.8,>=0.5 in /usr/local/lib/python3.6/dist-packages (from pytest>=2.0->pdfminer3k) (0.7.1)
Building wheels for collected packages: pdfminer3k
Building wheel for pdfminer3k (setup.py) ... done
Created wheel for pdfminer3k: filename=pdfminer3k-1.3.1-cp36-none-any.whl size=105356 sha256=a23e1d10fdda019d176374c394f8e83596c2f07a98a67
Stored in directory: /root/.cache/pip/wheels/ca/4f/a7/cb601b4fb257d2321ac668b7c6e269176780bd0283eda855d2
Successfully built pdfminer3k
Installing collected packages: ply, pdfminer3k
Successfully installed pdfminer3k-1.3.1 ply-3.11
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleu
```

```
Enter your authorization code:
.....
Mounted at /content/drive
/content/drive/My Drive/Peng
WARNING:root:Literal required: 'The process that creates this PDF constitutes a trade secret of codeMantra, LLC and is protected by the copy
WARNING:root:Literal required: '3-Heights(TM) PDF Producer 4.2.26.0 (http://www.pdf-tools.com)'
WARNING:root:Literal required: 'Quite Imposing 1.5d (EN)'
WARNING:root:Literal required: 'American Psychiatric Association'
WARNING:root:Literal required: LLC
['DIAGNOSTIC AND STATISTICAL \n', 'MANUAL OF\n', 'MENTAL DISORDERS\n', 'F I F T H E D I T I O N\n', 'DSM-5"\n', 'American Psychiatric Associ
['DIAGNOSTIC AND STATISTICAL \n', 'MANUAL OF\n', 'MENTAL DISORDERS', 'F I F T H E D I T I O N', 'DSM-5"', 'American Psychiatric Association',
'newtest.txt'
```

finish

Finish

The image shows the Google Drive web interface. On the left is a sidebar with navigation options: 'New', 'Priority', 'My Drive', 'Shared with me', 'Recent', 'Starred', 'Trash', and 'Storage' (28.6 GB used). The main area shows the 'My Drive' view with a breadcrumb 'My Drive > Peng'. A search bar is at the top. Below the breadcrumb is a table of files:

Name	Owner	Last modified	File size
DiagnosticandstatisticalmanualofmentaldisordersDS...	me	6:05 PM me	7 MB
newtest.txt	me	7:38 PM me	245 KB
Untitled0.ipynb	me	7:36 PM me	4 KB

The file 'newtest.txt' is highlighted with a blue background and circled in red. Below the table, the text 'Now the ouput file appear in your drive' is written in red.