

Sabry Saber <ssaber@minapharm.com>

Wed 4/6/2022 09:16 PM

To: Sabry Saber <ssaber@minapharm.com>

## Minapharm ML Engineer Technical Assesment:

### Objective:

You are provided with a medical related dataset, containing 5k articles. You are required to demonstrate your NLP, data-mining & topic extraction skills to extract the topics that are included within the dataset along with mapping each article to the corresponding topic names. Each document should be matched to the top three most relevant topic names including the matching probability score.

Kindly, mention how you came up with the optimum number of topics extracted and a final evaluation statement of your work.

### Requirements:

- It is recommended to use Python/R + LDA for topic name extraction. If you use other technology, you must mention your reasoning for your choice.
- Document your code and communicate your overall evaluation of the problem at hand
- You are expected to include along with your solution a report explaining:
  - Tools used
  - All performed steps, reasoning and outcomes
  - How you can further improve the results
  - How to further clean & optimize the data-sets to improve the performance of the LDA algorithm
  - Metrics used or to use if any to measure/ improve quality of results?

Dataset

<https://www.minapharm.com/gShare/Pubmed5k.rar>

**Deadline:** You have to send your work to [mena.abdelkodous@minapharm.com](mailto:mena.abdelkodous@minapharm.com) with email subject "ML Task 2022" before **Thursday April 14, 2022**

This e-mail has been sent to [ssaber@minapharm.com](mailto:ssaber@minapharm.com), [click here to unsubscribe](#).

Minapharm 2T ElBardissi St Takseem Asmaa Fahmy Ardel Golf Nasr City 11341 Cairo EG

Verified by Minapharm Development Team .

Best regards,

**Sabry Saber**

Minapharm Pharmaceuticals

IT Manager

Information Technology Department

Phone: (+2055) 412158/ 413115

Fax: (+2055) 413109

[www.minapharm.com](http://www.minapharm.com).

[www.rhein-minapharm.com](http://www.rhein-minapharm.com).



This message and its attachments may contain confidential or privileged information that may be protected by law; they should not be distributed, used or copied without authorization. If you have received this email in error, please notify the sender and delete this message and its attachments. As emails may be altered, Minapharm pharmaceuticals/Rhein Minapharm Biogenetics is not liable for messages that have been modified, changed or falsified.

**[MINAPHARM Pharmaceuticals, S.A.E.](#)**