Large Language Models for Inventory Optimization

...

Aliasgar Zakir Merchant

Tasks

- Test if Large Language Models are able to generate bayesian networks.

- (Ongoing) Test if Large Language Models are able to generate proper results with correct reasoning given different types of prompting strategies:

    - Zero Shot Prompting

    - Few Shot Prompting

Questions Asked

1. What is Inventory Optimization?
2. Why is Inventory Optimization so important?
3. What are the 20 factors affecting inventory optimization?
4. Among these 20 factors are there any inter-dependent factors? List these and show their dependency.
5. From these 20 factors what are the most important factors that are critical for inventory optimization?
6. Generate a bayesian network using all of the above critical factors.
7. To the above network, can you add the interdependent relations or have you already taken them into consideration?
8. Can you generate a python code for the above network with matplotlib functionality?

Results - What is Inventory Optimization?

Similarity:

- Systematic approach to answering the question.
- Explained every point in detail.

Difference:

- BARD tries to provide an overall answer while GPT-3.5 provides a more focused answer.

Result - Why is Inventory Optimization so important?

Similarity:

- Highly similar answers provided.
- All the right points taken into consideration
- Proper explanation was provided for their answers.

Difference:

- GPT-3.5 provides more reasons of importance than BARD reaching a wider range of factors.

Result - What are the 20 factors affecting inventory optimization?

Similarity

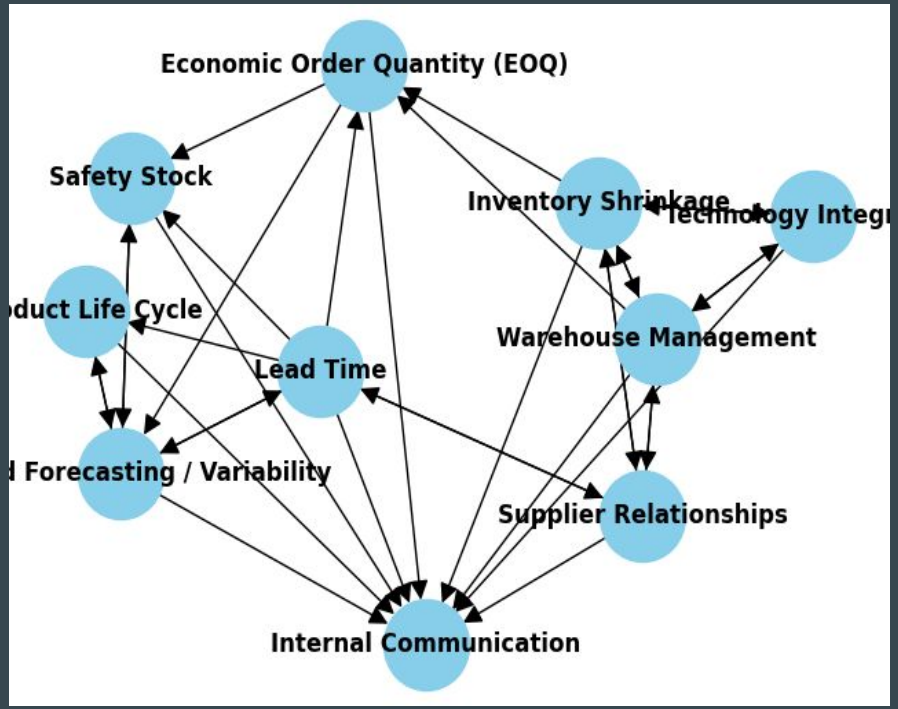
- Highly similar answers provided.
- Majority of the important points have been taken into consideration.
- Implicit relationships considered and I assume user is expected to know these.

Difference

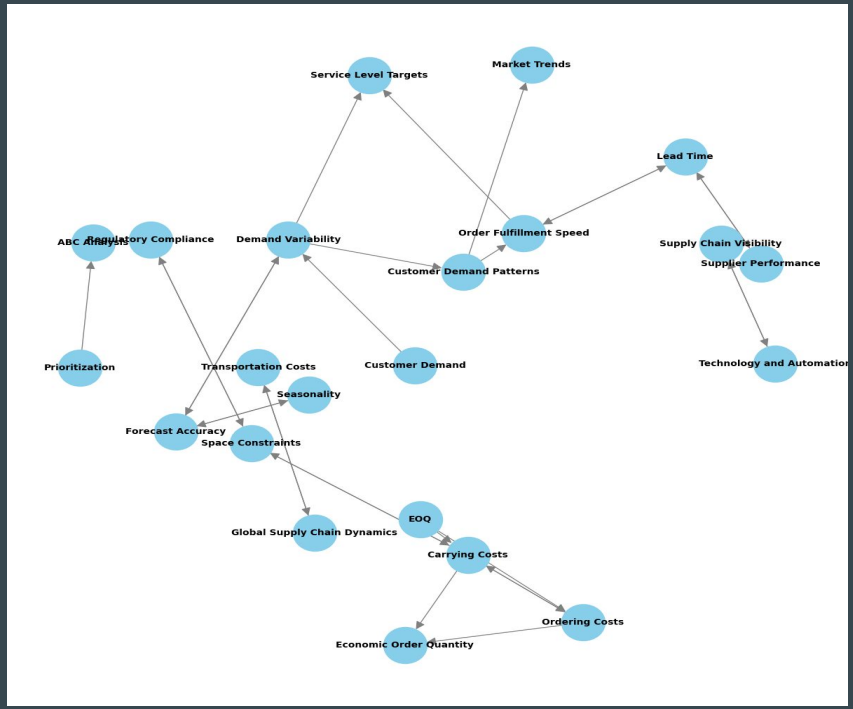
- BARD groups some factors together while GPT-3.5 separates them out.
- Answers vary with every try with a 20% variability factor for GPT-3.5 and 30-35% variability factor for BARD.

Result - Inter-dependent factors?

BARD



GPT-3.5



Result - Important factors critical for inventory optimization?

Similarity

- Similar answers generated with detailed explanation.
- Objectives mentioned by both models remain the same.

Difference

- While BARD generates an answer similar to GPT-3.5, the bayesian network and the factors are different to some extent.
- GPT-3.5 generates many important factors.
- While GPT-3.5 mentions that these factors can change based on real world data, BARD does not mention anything of such sort.

Result - Bayesian network using the above critical factors.

Similarity

- None of the models generated have CPT tables attributed to them.

Difference

- While BARD generates an answer similar to GPT-3.5, the bayesian network and the factors are different to some extent.

Result - Can you add the interdependent relations

Similarity

- BARD and GPT-3.5 mentions that they takes into account the interdependent relations.

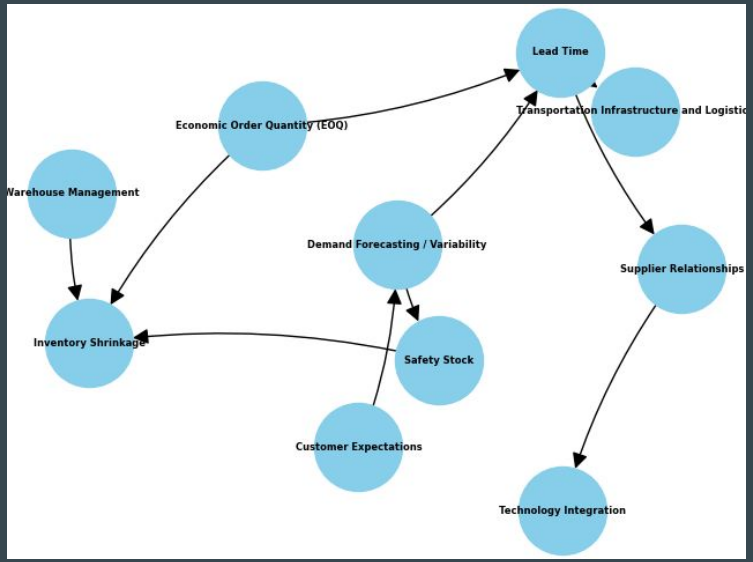
Difference

- BARD has some relations that might be implicitly declared or maybe even not declared in the bayesian network.
- GPT-3.5 on the other hand does specify that it takes into account the interdependent relationships and lists them correctly.
- GPT-3.5 mentions that the interdependencies captured are conceptual and may vary based on real world data and information while BARD does not.

Result - Python code with matplotlib functionality?

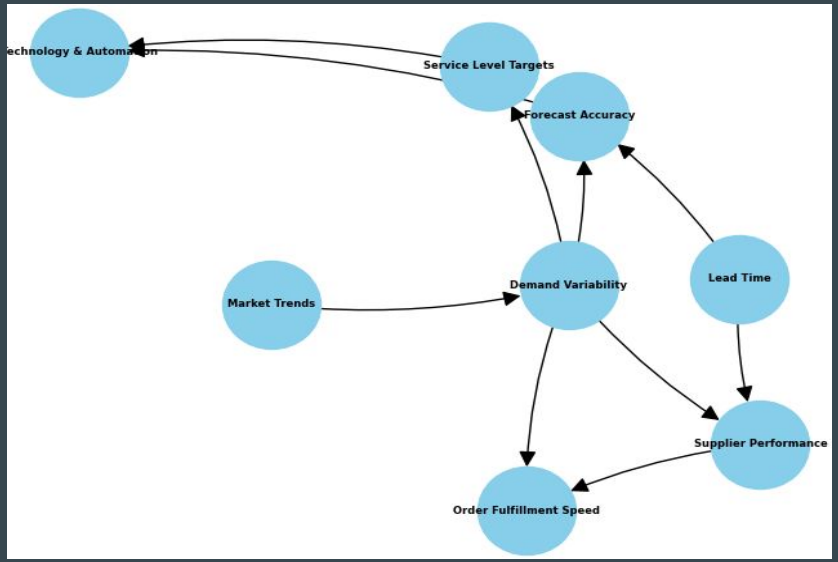
BARD

- Code does not work. Had to fix.
- Is not able to add visualization code. Had to add manually.



GPT-3.5

- Code works 90% of the time.
- Is able to add visualization code.



Additional Results

- BARD is good at understanding the user requirements and provides good answers, however unlike GPT-3.5 who remembers previous information to a certain extent, BARD is unable to do so. (Context: Different contextual queries)
- Both models generate very different results when the same query is run in another window, except when the API is used. (API Not Tested)
- Sometimes, the models generated dependencies involving self loops and produced code which did not accommodate self loops and did not know how to debug.
- Upon attempting to generate CPT tables, both models declare that the information is based on their understanding of the factors and that they may vary according to the real world data.

THANK YOU !!