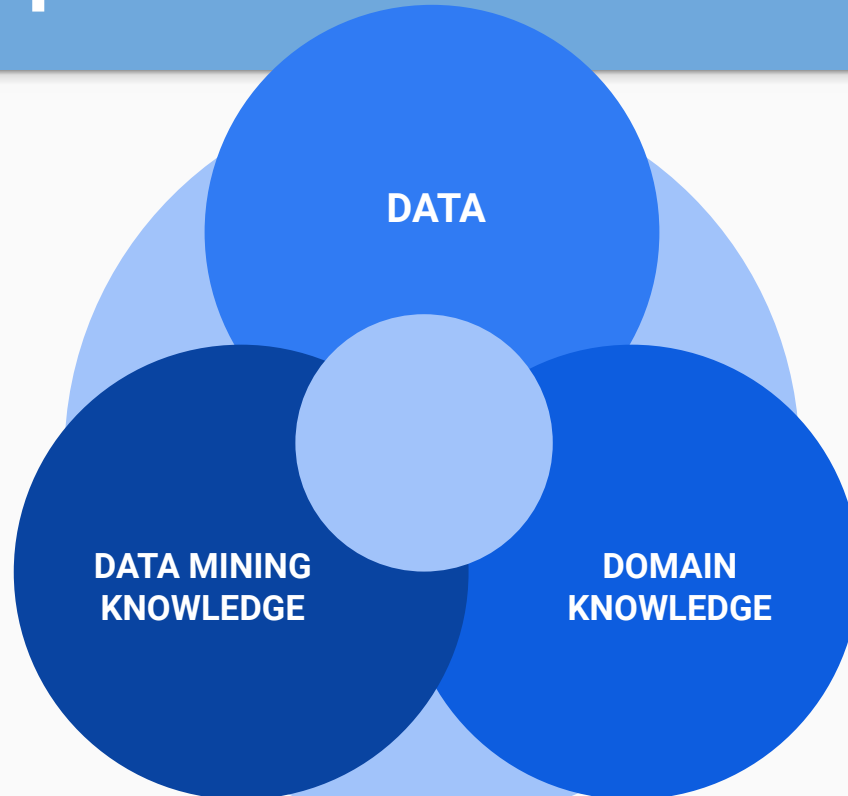




Quick Tips and Trick

Data Mining - Pra Gemastik XIII
25 July - Christian W. (IF 16)

What To Prepare?



Data

Publicly Available

1. <https://www.kaggle.com/datasets>
2. <https://data.bmkg.go.id/>
3. <https://archive.ics.uci.edu/ml/index.php>
4. <https://developer.twitter.com/en/docs>

kaggle™



BMKG



Domain Knowledge

1. **Framing problem** → real problem that can be solved using Data Mining Approach
2. **Stakeholder** → Who's going to use your proposal, how your proposal can help them

Tips:

1. Research **previous finalist** paper title
2. Read news to gather **emerging issue**

Data Mining Knowledge

Must Know Concept

1. **Machine Learning Task** : Classification, Regression, Sequence Generation
2. **Dataset Type** : Tabular, Time-series, Image, Text
3. **Feature Engineering and Selection**
4. **Model Selection**
 - a. **Model Type**: Tree-based, Linear, Deep Learning
 - b. **Evaluation Type**: cross-validation, hold-out
5. **Metrics**
 - a. **Regression**: RMSE, MAE, etc
 - b. **Classification**: AUC, Precision, Recall, etc

Low Resource?

The logo for Google Colab, featuring the word "colab" in a bold, lowercase, sans-serif font. The letters are orange, with the "c" and "o" having a slight gradient or shadow effect.The logo for Kaggle, featuring the word "kaggle" in a lowercase, sans-serif font. The letters are blue, with a small trademark symbol (TM) at the end of the word.