

Data scientist checklist for ML model collaborations

Prepared by Ashley Suh, Aug 2021

Before the modeling process starts:

- Outline or formalize the problem statement – this should fit on a single slide.
- Thoroughly establish with SMEs / collaborators:
 - What are the current practices and what will this model solve?
 - What strengths and what weaknesses are most crucial for choosing / developing this model?
 - Agree on an acceptance criteria before the modeling process begins – at the very least, model “must haves” or “would be nice to have” provided by SMEs
- Develop a common language for data scientists + SMEs to evaluate the performance of the model together:
 - The model can be evaluated based on the domain and the needs of the business
 - Decide on the foundation of the data science problem you’re solving to develop a set of metrics
 - Not all SMEs will know what they want ahead of time, so prepare a “model catalog” with strengths/weaknesses

During the development of a model:

- Prepare common documentation, establish consistent communication and check-ins (on-line / off-line), and send presentations ahead of time
- Explicitly and upfront state any errors, uncertainty, or limitations of the model to SME; identify outliers and debug models collaboratively with SMEs
- Establish a way for SMEs to provide context or feedback to the model (filling in missing data, identifying important features, correcting outputs, etc.)

When presenting on or visualizing performance of ML models:

- Make your presentation/tool accessible, with a descriptive prompt, to let SMEs get comfortable with it before the meeting
- Provide a “cheat sheet” for any potentially confusing vocabulary or language in your presentation; annotate plots & presentations with context and/or intended conclusions to draw
- Use interaction and live demos whenever appropriate.
 - Interactive plotting libraries and widgets for computational notebooks (such as jupyter, Observable, colab): ipywidgets, voila-dashboards, bqplot, plotly
 - Show SMEs: how the model input affects its output, how the data is distributed, how the model could be improved, where the weaknesses and strengths are (for data, performance)

Data scientist checklist for ML model collaborations

Prepared by Ashley Suh, Aug 2021

- Visually explain the significance and context of performance metrics (e.g., MAE, MSE) in your charts, while avoiding overly complex charts (audience most familiar with BBC infographics or NYT graphics)
- Provide a range of comparisons when communicating results of your model (e.g., compare to a baseline, current practices, best/worst case scenarios, etc.)
- When articulating results, start slow and offer to speed up. Be mindful that some SMEs come with their own “baggage” to a problem and may not always ask clarifying questions unless prompted

Presentation suggestions based on audience and time

