7 Biggest Mistakes when Documenting AI/ML Projects

Proper AI documentation enables teams to ship and update models faster. It also ensures business impact and compliance.

vectice

Introduction

Machine learning model documentation is often an afterthought, but clear and up-to-date documentation is essential for effective model development, deployment, and maintenance. Neglecting documentation can lead to major headaches down the road.

Creating documentation is time-consuming and can feel unrewarding, especially when teams iterate on models rapidly. However, skimping on documentation now frequently leads to confusion and wasted effort later as projects progress. Undocumented models become black boxes, making it hard for new team members to understand model logic and training data. Lack of documentation also makes maintaining and updating models extremely difficult.

1. Documenting after the fact

The first big mistake people make with machine learning: **waiting until the end of the project to write the documentation.**

Documenting too late is a problem because data science is an iterative process that includes a many details and context. It's just impossible to remember everything. If you wait too long, you might forget critical details, the choices made, and inputs from other stakeholders that were key to your work on datasets and models.

Don't miss the train, document as you work

Impact of this bad practice:

Losing this information leads to mistakes and makes the process of documenting very tedious. data scientists end up spending more time looking for details on their past work than actually documenting it. This can make the documentation feel like a research project instead of an easy process. Data Scientists in this situation often struggle and are left wondering what they did and why they made certain choices. That's how they forget key context about their datasets and decisions.

How to solve this?

Updating your process and starting to document as the team work is the key. In regulated industries, this usually results in 2 months saved on the model to market. For non-regulated industries, it's still time saved and drastically increase documentation quality.

2. Limiting the scope of documentation to code documentation

The second big mistake data scientists make: **only focusing on documenting the code.**

Data science and AI are way more than just code. They include datasets, models, graphs, controls, and more. This makes AI very different from pure engineering: iterating, involving many stakeholders and many parts.

Impact of this bad practice:

When data scientists limit their documentation to the code, it immediately impacts your documentation readability and scope. With a scope reduced to the code, it becomes challenging to maintain, upgrade models, and implement any governance.

AI is more than just code, open the scope!

How to solve this?

Good documentation should cover everything from the start of your project to the end. This includes preparing and cleaning data, making models, and more. Remember to include all parts of your project, not just the code. Holistic AI documentation is key!

3. Documenting for the wrong audience

The third big mistake in machine learning documentation is not considering your audience.

Many stakeholders are involved in AI projects, like other teams, managers, and even governance, legal and risk groups. Everyone needs to understand the work data scientists are producing.

How to solve this?

To fix this mistake, write easy-to-understand documents for everyone in the company. But also make sure to add enough details for the technical experts. It helps to have different versions or sections for different readers.

Lastly, talk the with stakeholders involved to determine what they need from the documentation. This will help include the most important information for everyone; making your ML documentation more inclusive.

4. Not linking your key assets to your documentation

The fourth big mistake in machine learning documentation is: **not connecting the main assets to your documentation.**

It's important to have easy access to the key information, like important datasets, model versions, graphs and code. This helps anyone understand the project without bothering the person who made it. It also enables always to keep the main components handy if anyone needs to dig deeper.

Don't lose the keys to your work!

How to solve this?

In your documentation, make sure to include links to all the important assets and make sure you link the right version! This makes it easy for you or someone else to continue working on the project or answer questions. Keep track of where everything is, like code locations, graphs, illustrations, and dataset versions. This will save you a lot of time.

5. Documenting alone in a bubble

Never document in a bubble

The fifth big mistake in machine learning documentation is **documenting alone and keeping it for yourself.**

It's important to have someone else check your work. Reviews can be quick and easy, but they help ensure your writing is clear and you don't forget anything important. One of the main goal of documentation is to secure the maintainability and the upgradability of your work. It can also inspire your teammates!

How to solve this?

Having another pair of eyes go through your documentation is very helpful, and a peer review process is a great way to do this.

6. Inconsistent documentation

The sixth big mistake in machine learning is having everyone in your team document without guidance.

When there are no guidelines, rules, or best practices for writing documentation, each data scientist will do it differently. Some will waste time over-documenting, while others will produce usable documentation.

As documentation stacks up, consistency is key!

How to solve this?

A good start is to **agree on simple guidelines for everyone to follow and tie them up in different phases of the life cycle.** This will help keep things clear easy to understand and ensure that the most important parts of the documentation are covered.

7. Manual documentation in the automation era

Time to stop manual documentation

The biggest mistake in machine learning documentation is: **document manually.**

This means switching from manual to automatic documentation. There are tools out there that help you easily capture and document your work as it happens. These tools also document all the parts related to data science, like data sets, models, codes, versions, and pictures, without any extra work.

Nowadays, automation tools can produce excellent first drafts of documentation in plain english, integrating details on your assets wisely and including key graphs. These tools can also make documentation for different people automatically. They keep track of all the parts and links as they are made and connect them to the documentation. Plus, they have a way for others to review your documentation and set rules.

How to solve this?

The most important thing is to **upgrade from manual to auto-documentation**. Doing this can make the challenging, tedious, and mistake-filled process of writing documentation a thing of the past.

Avoid Documentation Pitfalls by Embracing Automation

Machine learning documentation is challenging due to its time-consuming and error-prone nature. Neglecting proper documentation can lead to teamwork issues, model trust concerns, and regulatory compliance problems.

The 7 major errors to avoid are:

1. Documenting post-project completion
2. Focusing only on code
3. Ignoring diverse audience needs
4. Failing to link key assets
5. Documenting in isolation
6. Lack of standardized documentation practices
7. Sticking to manual documentation

With 🎲 vectice :

➔ Auto-documents as teams are working.

➔ Entire AI project auto-documented and works with Git.

➔ Generates custom documentation for your different audiences.

➔ Auto-links Datasets, Models, Code, Graphs and tracks lineage.

➔ Features a lightweight peer review system.

➔ Enabling to define and share best practices.

➔ Vectice auto-documentation boosts productivity and moral.

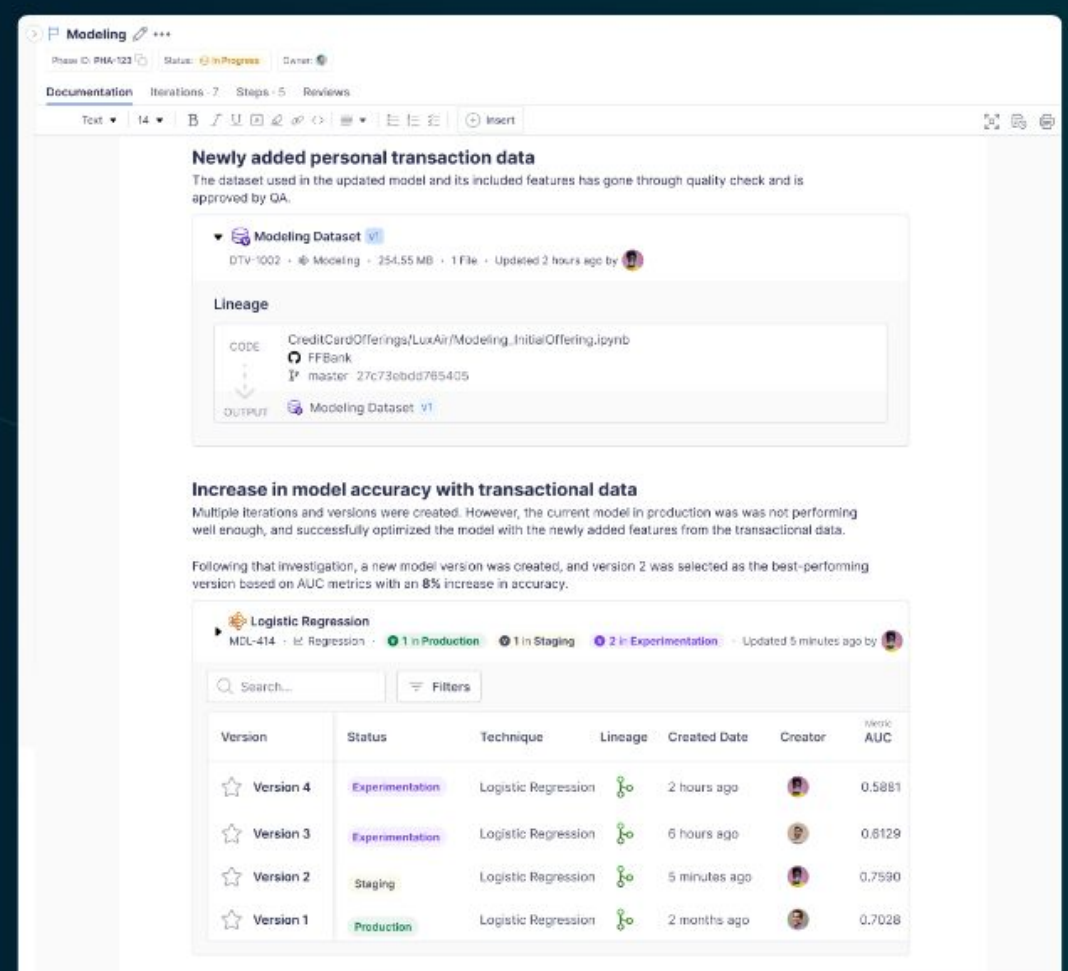Proper AI documentation enables teams to ship and update models faster while ensuring business impact and compliance.

Vectice is the Auto-Documentation Software for ML projects and their governance.

With Vectice, your team can effortlessly attain the documentation standards required for maintainability, explainability, and compliance with the ever-evolving AI regulatory landscape – all without the burden of error-prone manual labor.

With Vectice's unique technical and cross functional Auto-Documentation capabilities, you will accelerate your ML projects, delivering value faster with peace of mind.
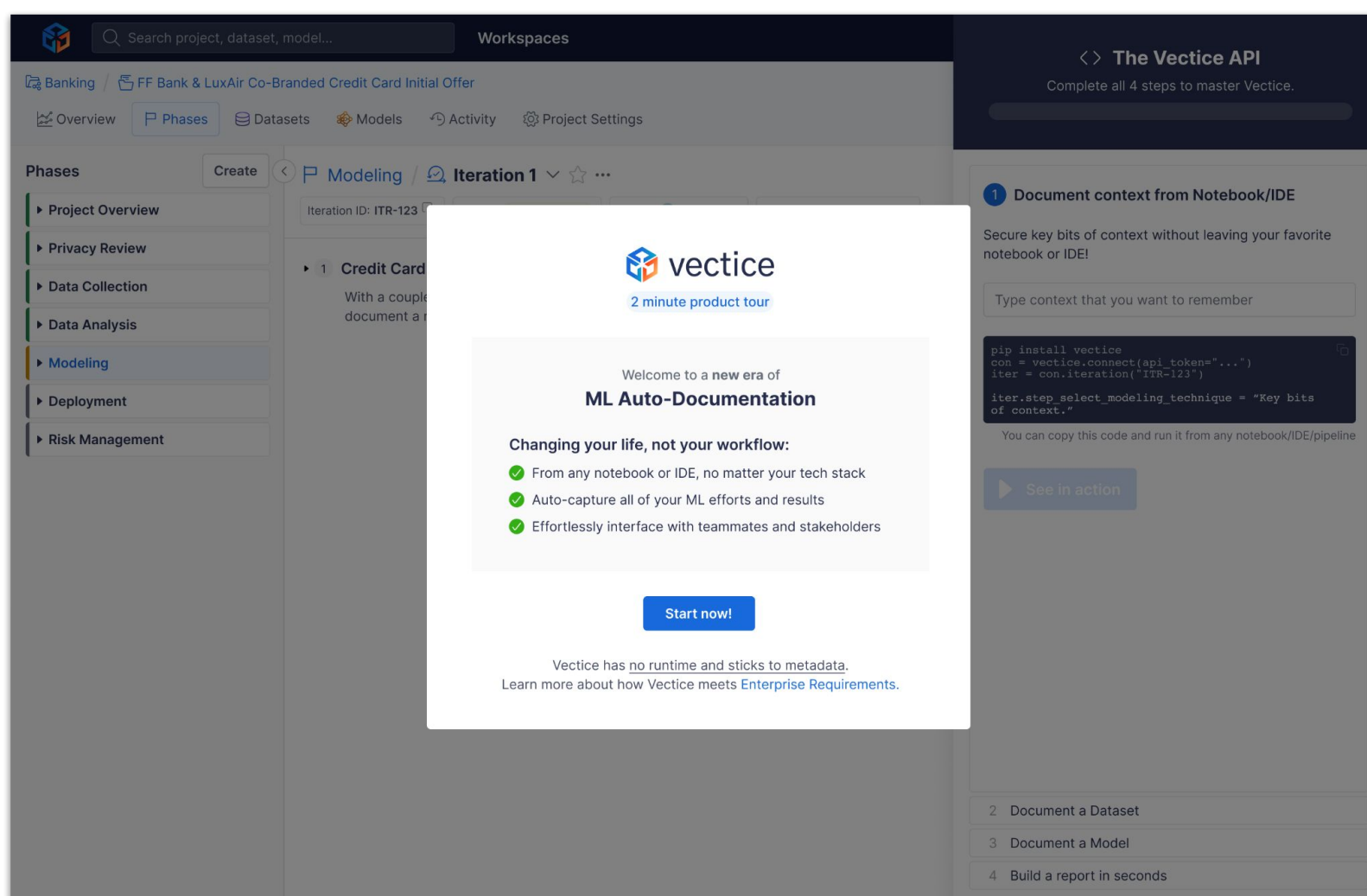
Try Vectice for free - www.vectice.com