# PyTorch Neural Network Model - One Pager

Fionn McGoldrick

September 17th 2025

## 1 Problem

Choosing the right programming language and framework is critical for any software project. Developers, especially beginners, often find themselves confused and wasting time in making this decision. This uncertainty can lead to making poor technology choices, slower development, and increased maintenance cost.

## 2 Audience

The primary audience is beginners to intermediate programmers who struggle with choosing the appropriate technologies for their desired project. This tool can support software students, interns, small teams, and IT consultants who need quick, evidence-based recommendations.

## 3 Solution

**Approach**

- *Data*: `CSV` file of labeled project briefs.
- *Data Loading*: Python script using `pandas` to read the CSV and structure it into labeled rows and columns.
- *Preprocessing*: Python script with `scikit-learn` and `NumPy` to tokenize and vectorize the dataset.
- *Training*: Python module with `PyTorch` to learn model weights from the processed dataset.
- *Exporting*: Script that converts the trained `PyTorch` model into an `ONNX` format for portable inference.
- *Serving*: Backend service using `FastAPI` to expose a `/predict` endpoint. It applies the preprocessing pipeline to incoming JSON and runs inference with the ONNX model.
- *Deployment*: Docker image running on `uvicorn`, with CI to build and test. An optional secondary container reproduces training.

## 4 Success Criteria

The system should achieve consistent accuracy with held-out test data, deliver low latency predictions on a standard CPU container, and remain fully reproducible using fixed seeds. It will be considered successful if the model can be retrained, exported, and deployed in Docker with minimal setup and clear documentation.

## 5 Out of Scope

The project does not aim to generate full application codebases, provide enterprise-grade security, or include real-time updates from external sources. Personalized recommendations requiring private organizational data are excluded.