CDW Documentation

Azure ML Pipeline Test

Azure ML Pipeline Test

Purpose

Test deployment of Azure ML Pipeline and look at outputs and download trained model.

Key Things Learned

1. When you create train.py and prep.py or any other environment files, they should be stored under ./src in your notebook directory.
2. You need to understand a bit of what you are trying to have it do as it can't think or make suppositions in a normal way. Garbage in, garbage out.

Final Code

train.py

```
import pandas as pd
import argparse
import os
from sklearn.linear_model import LogisticRegression
import joblib

def main():
    parser = argparse.ArgumentParser()
    parser.add_argument("--training_data", type=str)
    parser.add_argument("--model_output", type=str)
    args = parser.parse_args()

    df = pd.read_csv(os.path.join(args.training_data, "prepped.csv"))
    X = df[["feature1", "feature2", "feature_sum"]]
    y = df["label"]

    model = LogisticRegression()
    model.fit(X, y)

    os.makedirs(args.model_output, exist_ok=True)
    joblib.dump(model, os.path.join(args.model_output, "model.joblib"))

if __name__ == "__main__":
    main()
print("Model output path:", args.model_output)
print("Directory contents after writing:")
print(os.listdir(args.model_output))
print("Writing model to:", args.model_output)
```

```
print("Files in output dir:", os.listdir(args.model_output))
```

```
NOTE:  The print statements on the end were for troubleshooting and
shouldn't be there for production runs.
```

prep.py

```
import pandas as pd
import argparse
import os

def main():
    parser = argparse.ArgumentParser()
    parser.add_argument("--input_data", type=str)
    parser.add_argument("--output_data", type=str)
    args = parser.parse_args()

    df = pd.read_csv(args.input_data)
    df["feature_sum"] = df["feature1"] + df["feature2"]
    os.makedirs(args.output_data, exist_ok=True)
    df.to_csv(os.path.join(args.output_data, "prepped.csv"), index=False)

if __name__ == "__main__":
    main()
```

deployment_script.py

```
NOTE:  This is ran from the Notebook, not from a python script.  At least
not without changes.
```

AI Knowledge