8.Prescription_Features

December 5, 2025

1 Prescriptions (CPOE) Feature Script

1.1 Description

This script processes Provider Order Entry (CPOE) data from `prescriptions.csv`. It captures the physician's *intent* to treat, regardless of whether the drug was successfully administered.

1.2 Clinical Justification for HAPI Research

- **Risk Stratification:** Physician orders for Vasopressors or Sedatives act as an early warning signal for HAPI risk, often appearing in the system hours before the drug is physically administered.
- **Comorbidity Burden:** A high number of distinct prescriptions (`num_distinct_drugs`) is a validated proxy for patient frailty and medical complexity, which correlates with reduced tissue tolerance.

1.3 Inputs & Outputs

- **Input:** `prescriptions.csv`
- **Output:** `prescriptions_feat.csv`
- **Key Features:**
  - `has_vasoactive_drug`
  - `has_sedation_drug`
  - `num_prescriptions`

[7]:
```
import os
import pandas as pd
```

[8]:
```
#Configuration
import os
import pandas as pd

# Base Directory
BASE_DIR = r"D:\School\5141"

# Define Input/Output Paths
PRESC_PATH  = os.path.join(BASE_DIR, "prescriptions.csv", "prescriptions.csv")
OUTPUT_PATH = os.path.join(BASE_DIR, "prescriptions_feat.csv")
```

[9]:
```
# High-risk medication classes.

# GENERAL HIGH RISK
# Combined list of all drugs known to impact perfusion or mobility.
HIGH_RISK_MED_KEYWORDS = [
    "norepinephrine", "levophed", "epinephrine", "vasopressin",
    "dopamine", "dobutamine", "phenylephrine", "neosynephrine",
    "propofol", "midazolam", "versed", "lorazepam", "ativan",
    "dexmedetomidine", "precedex",
    "fentanyl", "morphine", "hydromorphone", "dilaudid", "oxycodone"
]

# VASOACTIVE AGENTS
# Peripheral Vasoconstriction leads to Ischemia.
VASOACTIVE_KEYWORDS = [
    "norepinephrine", "levophed", "epinephrine", "vasopressin",
    "dopamine", "dobutamine", "phenylephrine", "neosynephrine"
]

# SEDATIVES
# Immobility & Sensory Loss: Patients on these drips cannot move to relieve␣
 ↪pressure.
SEDATION_KEYWORDS = [
    "propofol", "midazolam", "versed", "lorazepam", "ativan",
    "dexmedetomidine", "precedex"
]

# OPIOIDS
# CNS Depression: Reduces spontaneous movement during sleep/rest.
OPIOID_KEYWORDS = [
    "fentanyl", "morphine", "hydromorphone", "dilaudid", "oxycodone"
]
```

[10]:
```
# Loading Function
def load_prescriptions(path: str):
    df = pd.read_csv(path, low_memory=False)

    if "hadm_id" not in df.columns:
        raise ValueError("prescriptions.csv is missing 'hadm_id'.")

    df["hadm_id"] = df["hadm_id"].astype("Int64")

    # Identify the best medication text columns
    possible_text_cols = [
        "drug",
        "drug_name_generic",
        "drug_name_poe",
```

```
        "medication"
    ]

    existing_cols = [c for c in possible_text_cols if c in df.columns]

    if not existing_cols:
        raise ValueError(
            f"No usable medication text column found. Columns: {list(df.
 ↪columns)}"
        )

    # Build combined free-text medication field
    df["med_text"] = (
        df[existing_cols]
        .astype(str)
        .agg(" ".join, axis=1)
        .str.lower()
    )

    return df
```

[11]:
```
# Building Features Function
def build_prescription_features(df: pd.DataFrame):

    df = df[df["hadm_id"].notna()].copy()

    # Count total prescriptions per admission
    num_prescriptions = (
        df.groupby("hadm_id")["med_text"]
        .size()
        .rename("num_prescriptions")
    )

    # Count distinct meds per admission
    num_distinct = (
        df.groupby("hadm_id")["med_text"]
        .nunique()
        .rename("num_distinct_drugs")
    )

    # Helper for keyword-based counts/flags
    def count_and_flag(keywords, count_name, flag_name):
        pattern = "|".join(keywords)
        mask = df["med_text"].str.contains(pattern, na=False)

        count_series = (
            df[mask]
```

```
            .groupby("hadm_id")["med_text"]
            .size()
            .rename(count_name)
        )

        flag_series = (
            df[mask]
            .groupby("hadm_id")
            .size()
            .gt(0)
            .astype("Int64")
            .rename(flag_name)
        )

        return count_series, flag_series

    # High-risk categories
    high_count, high_flag = count_and_flag(
        HIGH_RISK_MED_KEYWORDS,
        "num_high_risk_drugs",
        "has_high_risk_drug"
    )

    _, vaso_flag = count_and_flag(
        VASOACTIVE_KEYWORDS,
        "num_vasoactive_drugs",
        "has_vasoactive_drug"
    )

    _, sed_flag = count_and_flag(
        SEDATION_KEYWORDS,
        "num_sedation_drugs",
        "has_sedation_drug"
    )

    _, opioid_flag = count_and_flag(
        OPIOID_KEYWORDS,
        "num_opioid_drugs",
        "has_opioid_drug"
    )

    # Combine into a single feature table
    feat = (
        num_prescriptions.to_frame()
        .join(num_distinct, how="left")
        .join(high_count, how="left")
        .join(high_flag, how="left")
```

```
        .join(vaso_flag, how="left")
        .join(sed_flag, how="left")
        .join(opioid_flag, how="left")
        .reset_index()
        .fillna(0)
    )

    # Ensure numeric columns are Int64
    int_cols = [
        "num_prescriptions",
        "num_distinct_drugs",
        "num_high_risk_drugs",
        "has_high_risk_drug",
        "has_vasoactive_drug",
        "has_sedation_drug",
        "has_opioid_drug"
    ]

    for col in int_cols:
        if col in feat.columns:
            feat[col] = feat[col].astype("Int64")

    print("Prescription features built:")
    print(feat.head())

    return feat
```

[12]:
```
# Execute

if __name__ == "__main__":
    df = load_prescriptions(PRESC_PATH)
    feat = build_prescription_features(df)

    print(f"Saving to: {OUTPUT_PATH}")
    feat.to_csv(OUTPUT_PATH, index=False)
    print("Done.")
```

```
Prescription features built:
    hadm_id  num_prescriptions  num_distinct_drugs  num_high_risk_drugs  \
0  20000019                 29                  18                    0
1  20000024                 17                   9                    4
2  20000034                 37                  28                    1
3  20000041                 34                  27                    3
4  20000045                105                  31                   13

   has_high_risk_drug  has_vasoactive_drug  has_sedation_drug  has_opioid_drug
0                   0                    0                  0                0
1                   1                    0                  0                1
```

```
2                   1                   0                   0                   1
3                   1                   0                   0                   1
4                   1                   0                   0                   1
Saving to: D:\School\5141\prescriptions_feat.csv
Done.
```