

# Johns Hopkins AI Hackathon

## Data Documentation

December 1-2, 2023

# Introduction

Protenus is a Baltimore-based artificial intelligence company that helps health systems automate compliance and auditing to drive the safest patient outcomes while protecting organizational reputation. Protenus is committed to innovation, determined to reduce risk, and focused on supporting our community of employees, customers, and ultimately, patients.

Founded in 2014, Protenus has been recognized by Gartner as a Cool Vendor in Healthcare Artificial Intelligence, by CB Insights as one of the Most Innovative Digital Health Startups, by Forbes as one of America's Top Startup Employers, and by Modern Healthcare as one of the Best Places to Work in Healthcare.

Protenus sells several AI-enabled products on the Protenus platform, including Privacy Monitoring and Diversion Monitoring. Privacy Monitoring detects privacy and security fraud such as HIPAA violations and snooping. Diversion Monitoring detects various types of drug theft that can contribute to the opioid epidemic.

Protenus sells to hospital systems in the United States. Our ideal customers are large health systems with several associated hospitals. During the sales process, Protenus tracks sales data, creating a monthly snapshot of each active sales opportunity associated with known account information.

## Definitions

### Basic Terms

**Account:** A prospective or current customer account. In this case, accounts consist of health system enterprises, usually made up of many hospitals. For example, the Johns Hopkins Health System is an account made up of 5+ hospitals. Account data is exported from an external data source called [Definitive Healthcare](#) that aggregates data from publicly available statements. Not all accounts have data that is publicly available so that is why many of the opportunities are missing some account information.

**Opportunity:** An opportunity to sell one or more product to a customer Account. Each opportunity progresses month to month through a sales cycle with attributes that are updated by the sales team in the sales database.

### Training Data Elements

- A. **Snapshot Date (Date):** The date that the data was captured. Snapshots are captured monthly. There are a few months that the monthly snapshot did not capture data, so those months are missing from this data set.

- B. **Opportunity ID (String):** A unique identifier for each opportunity captured each snapshot until the opportunity is no longer active (closed won or closed lost)
- C. **Opportunity Stage (String):** The sequential stage of the opportunity in the sales cycle. Ideally, each opportunity moves through each stage in order. In reality, customer accounts may skip stages or reverse stages as communications between the customer account and the sales team progress. Active opportunities exist in one of five stages before they are closed as either won or lost.
- a. Stage 0 - Prospecting: An opportunity has been identified for an account.
  - b. Stage 1 - Evaluating: The account is engaged with the sales team. The prospective customer account begins evaluation Protensus' solution(s).
  - c. Stage 2 - Scoping: The sales team begins scoping the size and scope of the opportunity based on the customer accounts input.
  - d. Stage 3 - Pricing: The sales team presents a proposal to the customer account.
  - e. Stage 4 - Verbal / VOC: The prospective customer account verbally confirms that Protensus is the vendor of choice for the opportunity.
  - f. Stage 5 - Contracting: The customer account begins contract negotiations with Protensus
  - g. Closed Won: The opportunity is won and no longer active.
  - h. Closed Lost: The opportunity is lost and no longer active.
- D. **Opportunity Type (String):** When sales opportunities become available for a new customer account, it is classified as New Business. When an opportunity becomes available for an account that is already a customer, it is classified as Existing Business. Existing Business opportunities are ideally closed faster than New Business because the account is an existing customer, a relationship is already established, and a contract is already in place.
- E. **Opportunity Annual Amount (Float):** The annualized amount of recurring revenue associated with each opportunity. This amount may increase or more likely decrease before the opportunity is closed, typically due to discounts to win the opportunity.
- F. **Opportunity Created Date (Date):** The date that the opportunity is created in the sales database.
- G. **Opportunity Close Date (Date):** The date that the opportunity is closed or expected to close in the sales database. If the opportunity is Closed Lost, this is the date that the customer account informed the sales team that the opportunity was lost. If the opportunity was Closed Won, this is the date that the customer contract was executed. If the opportunity is still open, the closed date is set by the sales team as an estimate for when the opportunity will be closed.
- H. **Opportunity Product (String):** The product that is being sold in to the custom account in the sales opportunity. The two primary products are Diversion Monitoring (detection of pharmacy fraud or theft) and Privacy Monitoring (detection of privacy/security fraud). When Diversion and Privacy Monitoring are sold together, the Product is referred to as Compliance Analytics (a platform with more than one AI solution)
- I. **Account Electronic Medical Record System (String):** The electronic medical record (EMR) vendor for each opportunity. The major EMR vendors are Epic, Cerner, Meditech, but many others exist. Epic is the predominant EMR vendor, followed by

Cerner, then Meditech. A health system that purchases a certain EMR may do so for a variety of reasons, but the decision could be compared to why a consumer may select Mac, Windows, or Linux as a preferred OS.

- J. **Account Hospital Type (String):** They type of customer account, including Childrens Hospital, Health System, Critical Access Hospital, Short Term Acute Care Hospital, Psychiatric Hospital, Rehabilitation Hospital, or Long Term Acute Care Hospital.
- K. **Account Fiscal Year End (Date):** The date that the customer account uses as their fiscal year end. Any business can select when their fiscal calendar begins and ends. A fiscal year end may or may not impact the sales process depending on if software purchases are associated with the fiscal year budgetary process.
- L. **Account State (String):** The state in which the customer account is headquartered.
- M. **Account Zip Code (String):** The zip code in which the customer account is headquartered.
- N. **Account Number of Hospitals (String):** The total number of hospitals associated with the customer account.
- O. **Account Number of Beds (Int):** The total number of patient beds associated with the customer account.
- P. **Account Number of Affiliated Physicians (Int):** The number of physicians affiliated with the customer account.
- Q. **Account Number of Employee (Int):** The number of employees affiliated with the customer account. Not all workforce members are employees. Academic medical centers may have faculty in addition to staffed employees. Any customer account may also have students, contractors, or other non-employed workforce members.
- R. **Account Cash on Hand (Int):** The amount of cash that a customer account has available based on public reporting.
- S. **Account Total Patient Revenue (Int):** Total patient revenue is the total revenue generated from patient services.
- T. **Account Net Patient Revenue (Int):** Net patient revenue (NPR) is the aggregate money generated from patient services collected from payors, including private insurance, Medicaid and Medicare. The calculation for NPR is the total patient revenues minus patient discounts. If a healthcare organization has a negative NPR, its discounts are greater than its total revenue. NPR does not factor in contractual adjustments, charitable donations, foundation earnings or bad debt.