#RahoAmbitious

LJMU upGrad

Research Overview of Recent Trends in Predictive Analytics, Recommendation Systems and Forecasting

Dr. Sahil Sharma, PhD
SME @ upGrad

Today’s Agenda

1. **Predictive Analytics**

2. **Recommendation Systems**

3. **Forecasting**

Predictive Analytics

Where Predictive Analytics Stands?

upGrad

Data analytics—examining data to answer questions, identify trends, and extract insights—can provide you with the information necessary to strategize and make impactful business decisions.

There are four key types of data analytics:

- Descriptive, which answers the question, “What happened?”
- Diagnostic, which answers the question, “Why did this happen?”
- Prescriptive, which answers the question, “What should we do next?”
- **Predictive,** which answers the question, “What might happen in the future?”

Reference: What Is Predictive Analytics? 5 Examples | HBS Online

What is Predictive Analytics?

upGrad

- Predictive analytics is a branch of advanced analytics that predicts future outcomes using historical data combined with statistical modeling, data mining techniques, and machine learning.

- Companies employ predictive analytics to find patterns in this data to identify risks and opportunities. Predictive analytics is often associated with big data and data science.

Types of Predictive Modeling

upGrad

- **Classification Models:** Types of classification models include logistic regression, decision trees, random forest, neural networks, and Naïve Bayes, etc.

- **Clustering Models:** Common clustering algorithms include k-means clustering, mean-shift clustering, density-based spatial clustering of applications with noise (DBSCAN), expectation-maximization (EM) clustering using Gaussian Mixture Models (GMM), and hierarchical clustering.

- **Time Series Models:** Autoregressive (AR), moving average (MA), ARMA, and ARIMA models are all frequently used time series models.

Industrial Use Cases

upGrad

- **Banking:** Financial services use machine learning and quantitative tools to predict credit risk and detect fraud.
- **Healthcare:** Predictive analytics in health care is used to detect and manage the care of chronically ill patients and to track specific infections such as sepsis.
- **Human Resources:** HR teams use predictive analytics and employee survey metrics to match prospective job applicants, reduce employee turnover and increase employee engagement.
- **Marketing and Sales:** Predictive analytics enables companies to engage with their clients across the customer lifecycle proactively.
- **Supply Chain:** Businesses commonly use predictive analytics to manage product inventory and set pricing strategies.

Predictive Analytics Examples

Predictive analytics helps teams in industries as diverse as finance, healthcare, pharmaceuticals, automotive, aerospace, and manufacturing.

- **Automotive** – Breaking new ground with autonomous vehicles
  Companies developing driver assistance technology and new autonomous vehicles use predictive analytics to analyze sensor data from connected vehicles and to build driver assistance algorithms.
- **Aerospace** – Monitoring aircraft engine health
  To improve aircraft up-time and reduce maintenance costs, an engine manufacturer created a real-time analytics application to predict subsystem performance for oil, fuel, liftoff, mechanical health, and controls.
- **Energy Production** – Forecasting electricity price and demand
  Sophisticated forecasting apps use models that monitor plant availability, historical trends, seasonality, and weather.
- **Financial Services** – Developing credit risk models
  Financial institutions use machine learning techniques and quantitative tools to predict credit risk.
- **Industrial Automation and Machinery** – Predicting machine failures
  A plastic and thin film producer saves 50,000 Euros monthly using a health monitoring and predictive maintenance application that reduces downtime and minimizes waste.
- **Medical Devices** – Using pattern-detection algorithms to spot asthma and COPD
  An asthma management device records and analyzes patients' breathing sounds and provides instant feedback via a smart phone app to help patients manage asthma and COPD.

Benefits of Predictive Modeling

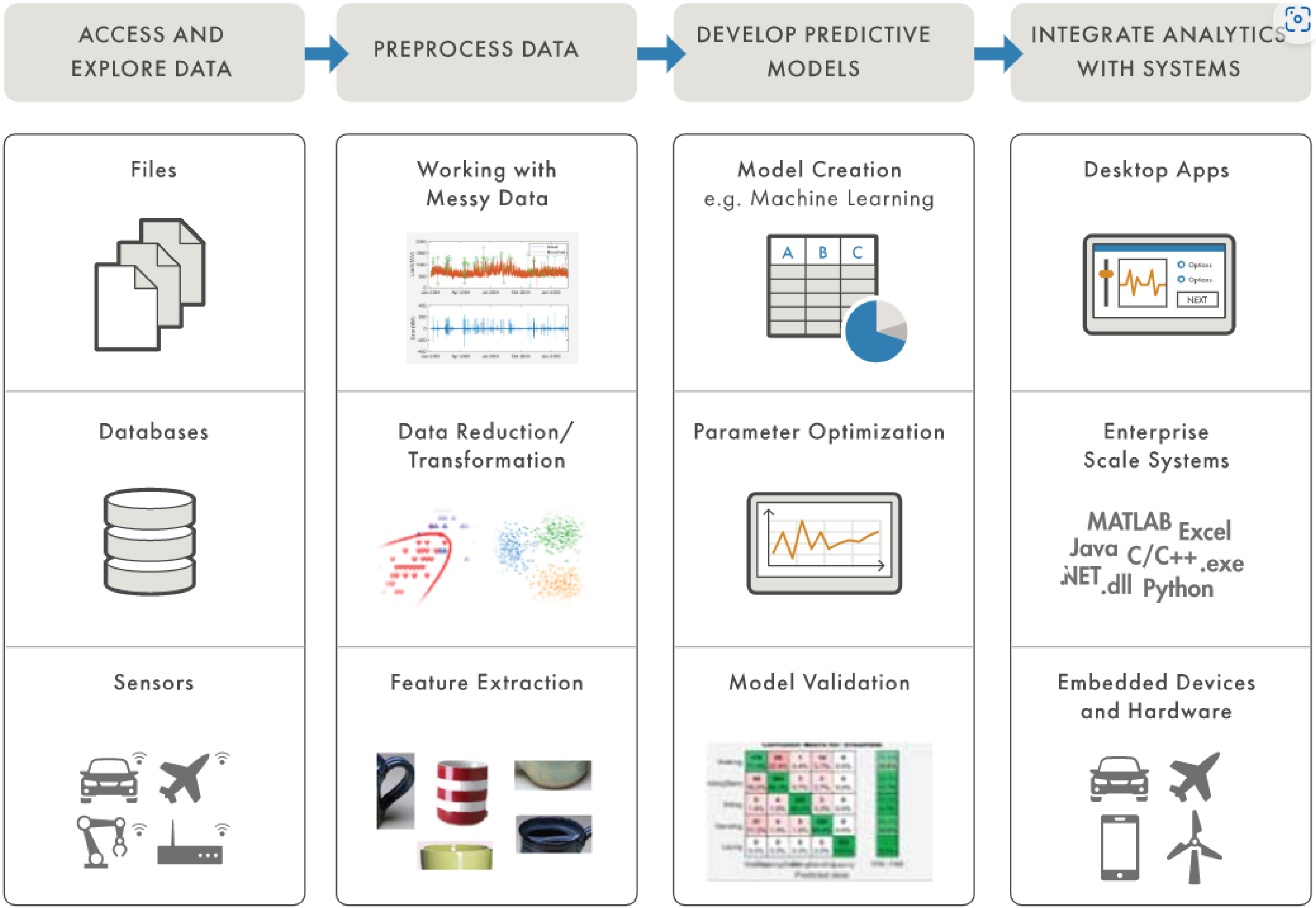
upGrad

- **Security:** Every modern organization must be concerned with keeping data secure. A combination of automation and predictive analytics improves security.
- **Risk Reduction:** In addition to keeping data secure, most businesses are working to reduce their risk profiles. For example, a company that extends credit can use data analytics to understand better if a customer poses a higher-than-average risk of defaulting.
- **Operational Efficiency:** More efficient workflows translate to improved profit margins.
- **Improved Decision Making:** Predictive analytics can provide insight to inform the decision-making process and offer a competitive advantage.

Predictive Analytics Workflow

What Is Predictive Analytics? - 3 Things You Need to Know - MATLAB & Simulink (mathworks.com)

upGrad



Predictive Analytics in Data Engineering

upGrad

1. Large-Scale Data Processing for Predictive Analytics

Research on how to efficiently process and manage large-scale datasets for predictive analytics. This could involve the use of distributed computing frameworks like Hadoop and Spark.

2. Real-Time Data Processing for Predictive Analytics

Research on real-time data processing frameworks and techniques for predictive analytics. This could involve streaming data technologies like Apache Kafka or Apache Flink.

3. Optimizing Data Pipelines for Predictive Analytics

Research on techniques for optimizing data pipelines to support predictive analytics, including data extraction, transformation, and loading (ETL) processes.

Fill Feedback Form 1

Recommendation Systems

What is a Recommendation System?

upGrad

A recommendation system is an artificial intelligence or AI algorithm, usually associated with machine learning, that uses Big Data to suggest or recommend additional products to consumers.

These can be based on various criteria, including past purchases, search history, demographic information, and other factors.

Recommender systems are beneficial as they help users discover products and services they might otherwise not have found.

Terminology for Recommendation Systems

upGrad

Items (also known as documents)

The entities a system recommends. For the Google Play store, the items are apps to install. For YouTube, the items are videos.

Query (also known as context)

The information a system uses to make recommendations. Queries can be a combination of the following:

- user information
  - the id of the user
  - items that users previously interacted with
- additional context
  - time of day
  - the user's device

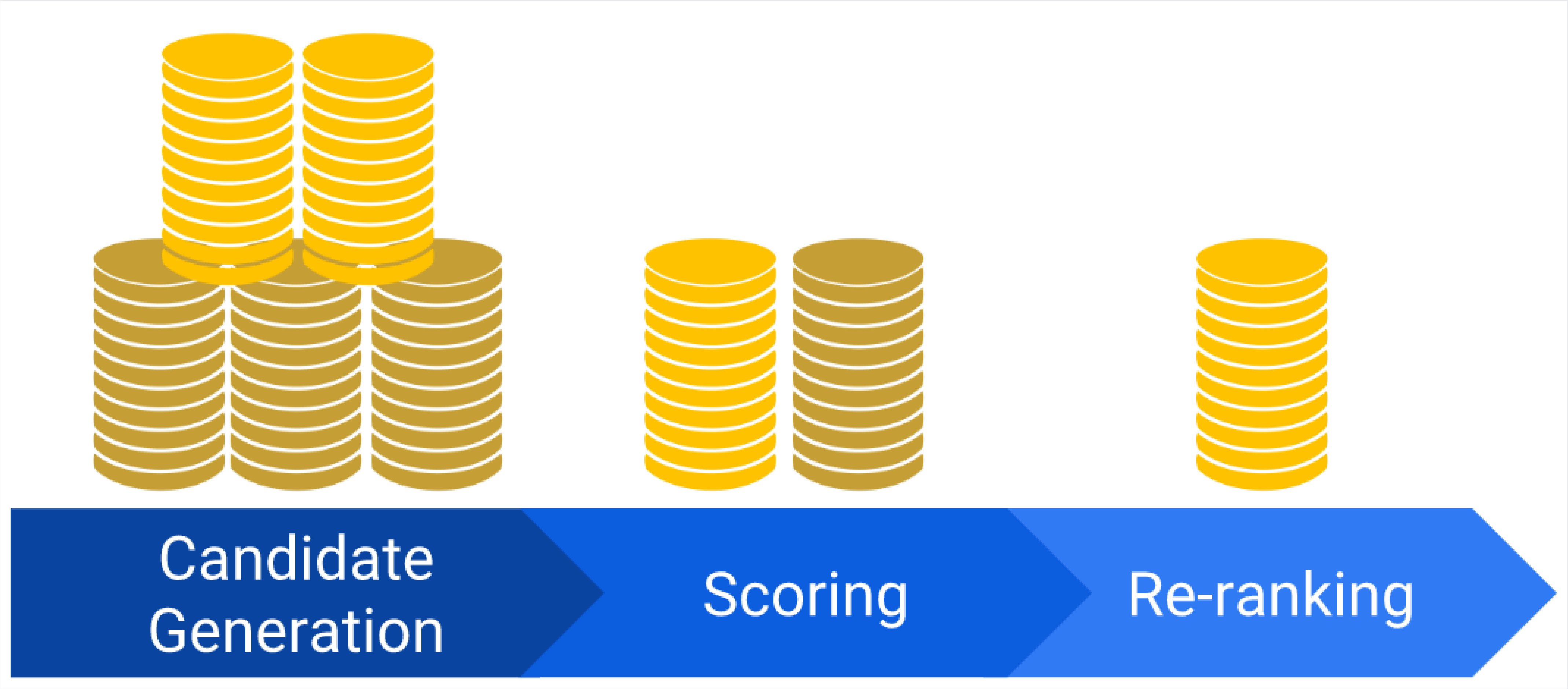
Embedding

A mapping from a discrete set (in this case, the set of queries or the set of items to recommend) to a vector space called the embedding space. Many recommendation systems rely on learning an appropriate embedding representation of the queries and items.

Recommendation Systems Overview

upGrad

One typical architecture for recommendation systems consists of the following components:



Recommendation Systems Overview – Contd.

upGrad

Candidate Generation

In this first stage, the system starts from a potentially huge corpus and generates a much smaller subset of candidates. For example, the candidate generator on YouTube reduces billions of videos to hundreds or thousands. The model needs to evaluate queries quickly, given the enormous size of the corpus. A given model may provide multiple candidate generators, each nominating a different subset of candidates.

Scoring

Next, another model scores and ranks the candidates to select the set of items (on the order of 10) to display to the user. Since this model evaluates a relatively small subset of things, the system can use a more precise model relying on additional queries.

Re-ranking

Finally, the system must consider additional constraints for the final ranking. For example, the system removes items the user explicitly dislikes or boosts the score of fresher content. Re-ranking can also help ensure diversity, freshness, and fairness.

Application of Recommendation Systems

upGrad

- E-Commerce & Retail: Personalized Merchandising
- Media & Entertainment: Personalized Content
- News Articles
- Personalized Banking
- Tag Recommendation

Collaborative vs Content Filtering

upGrad

- Collaborative Filtering:
  - The main idea that governs the collaborative methods is that through past user-item interactions, when processed through the system, it becomes sufficient to detect similar users or similar items to make predictions based on these estimated facts and insights.

- Content Filtering:
  - The main idea of content-based methods is to try to build a model, based on the available "features", that explain the observed user-item interactions. Still considering users and movies, we can also create a model to explain why so is happening.

Recommendation Systems in Data Engineering

upGrad

1. Data Engineering for Scalable Recommendation Systems: Investigation into data management techniques for handling the large, sparse datasets typically used in recommendation systems.

2. Real-Time Recommendation Systems: Research on how to design data engineering systems to support real-time recommendations, which require very fast data processing and updating.

3. Privacy-preserving Data Processing in Recommendation Systems: Research on techniques for processing and storing data in recommendation systems in a way that preserves user privacy.

Forecasting

What is a Forecast?

A forecast is a prediction made by studying historical data and past patterns. Businesses use software tools and systems to analyze large amounts of data collected over a long period. The software predicts future demand and trends to help companies make more accurate financial, marketing, and operational decisions.

Why is Forecasting Important?

upGrad

Forecasting acts as a planning tool to help enterprises prepare for the uncertainty that can occur in the future. It allows managers to respond confidently to changes, control business operations, and make strategic decisions that drive future growth. For example, businesses use forecasting to do the following:

- Use resources more efficiently
- Visualize business performance
- Time for the launch of new products or services
- Estimate recurring costs
- Predict future events like sales volumes and earnings
- Review management decisions

Types of Forecasting Methods

upGrad

Qualitative methods

Qualitative forecasting relies on marketing experts to make short-term predictions. You can use qualitative methods when there needs to be more historical data. For example, these are two use cases:

•Market research techniques like polls and surveys identify consumer demand.

•Delphi modeling techniques poll experts in a particular field to collect their opinions and predict trends in that field.

Quantitative methods

Quantitative forecasting models use meaningful statistics and historical data to predict long-term future trends. We give examples of standard quantitative methods below:

•Econometric modeling analyzes financial data sets, like loan and investment data, to predict significant economic shifts and their impact on the company.

•The indicator approach compares data points to identify relationships between seemingly unrelated data. For example, you can use changes in GDP to forecast unemployment rates.

•In this scenario, GDP data is called the lead indicator, and the unemployment rate is the lagging indicator.

•Time-series forecasting analyzes data collected over different intervals to predict future trends.

What is Time Series Data?

upGrad

- Cross-sectional data observes individuals and companies at the same time. On the other hand, time-series data is any data set that collects information at various intervals. This data is distinct because it orders data points by time. As a result, there is potential for correlation between observations in adjacent intervals.
- Time-series data is plottable on a graph with incremental intervals (or timelines) on the x-axis and observed sample data values on the y-axis. Such time-series graphs are valuable tools for visualizing the data. Data scientists use them to identify forecasting data characteristics.

Example of Time Series Data

upGrad

Time trending data

In trending data, y-values increase or decrease with time, making the graph appear linear. For example, population data may increase or decrease linearly with time.

Seasonality

Seasonal patterns occur when time-series data shows regular and predictable patterns at intervals of less than a year. This data pattern may appear as spikes or anomalies on a linear graph. For example, a store's retail sales might increase during the holidays around December and April.

Structural breaks

Sometimes time-series data suddenly changes behavior at a certain point in time. The time-series graph may suddenly shift up or down, creating a structural break or non-linearity. For instance, many economic indicators changed sharply in 2008 after the start of the global financial crisis.

Time Series Forecasting

upGrad

Time-series forecasting is a data science technique that uses machine learning and other computer technologies to study past observations and predict future values of time-series data. Let's look at some examples of time-series forecasting:

- Astronomical data consists of repetitive movements of the planets over centuries. You can use this data to accurately predict astronomical events like eclipses and comets.
- Weather forecasting uses wind and temperature patterns to predict weather changes.
- Scientists can use birth rates and migration data to predict population growth.

Time Series – Analysis vs Forecasting

Time-series analysis explores the underlying causes in any time-series data. This field seeks to understand the “why” behind a time-series dataset. Analysts must often make assumptions and decompose or break down the data to extract meaningful statistics and other characteristics.

While time-series analysis is about understanding the dataset, forecasting is about predicting it. These are the three steps of predictive modeling:

- Ask a question and collect a sample set of time-series data that answers this question for a past period.
- Train the computer software or forecasting algorithm using past values.
- Use the forecasting algorithm to make future observations.

Forecasting in Data Engineering

upGrad

1. Time-Series Database for Forecasting: Investigation into the design and use of time-series databases, which are specialized for handling time-series data commonly used in forecasting.

2. Data Engineering for High-Dimensional Forecasting: Research on data management techniques for high-dimensional forecasting problems, which involve forecasting multiple interrelated time series.

3. Handling Uncertainty in Forecasting Data Pipelines: Research on techniques for managing and propagating uncertainty in forecasting data pipelines.

Fill Feedback Form 2

Any Questions?

Thank you!

Let’s connect on LinkedIn:
https://www.linkedin.com/in/sahil301290/