

# TUTORIAL 1

## ORANGE, RAPIDMINER, AND DATA PREPROCESSING

---

Aut, CS228, 2026, NS

# Contents

---

- Gange
- Repchner
- Data Processing Examples





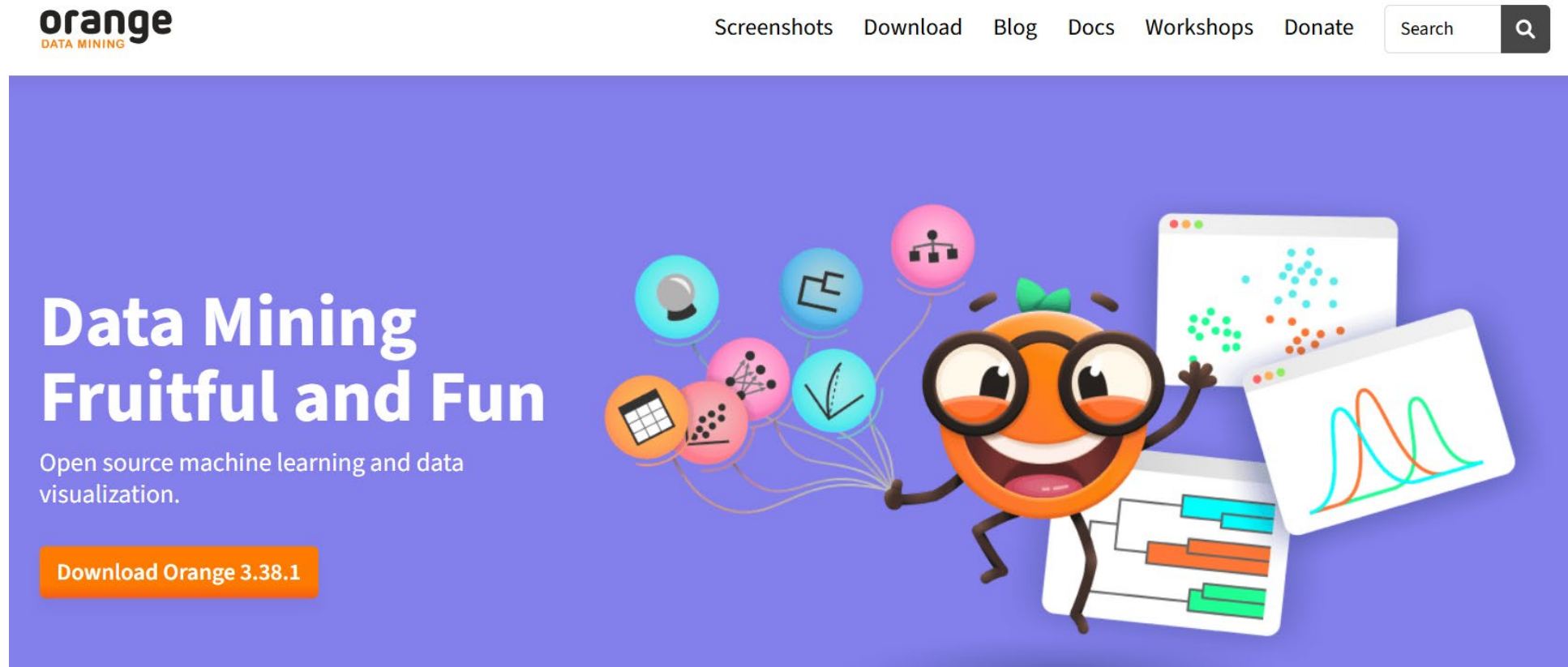
Orange

---



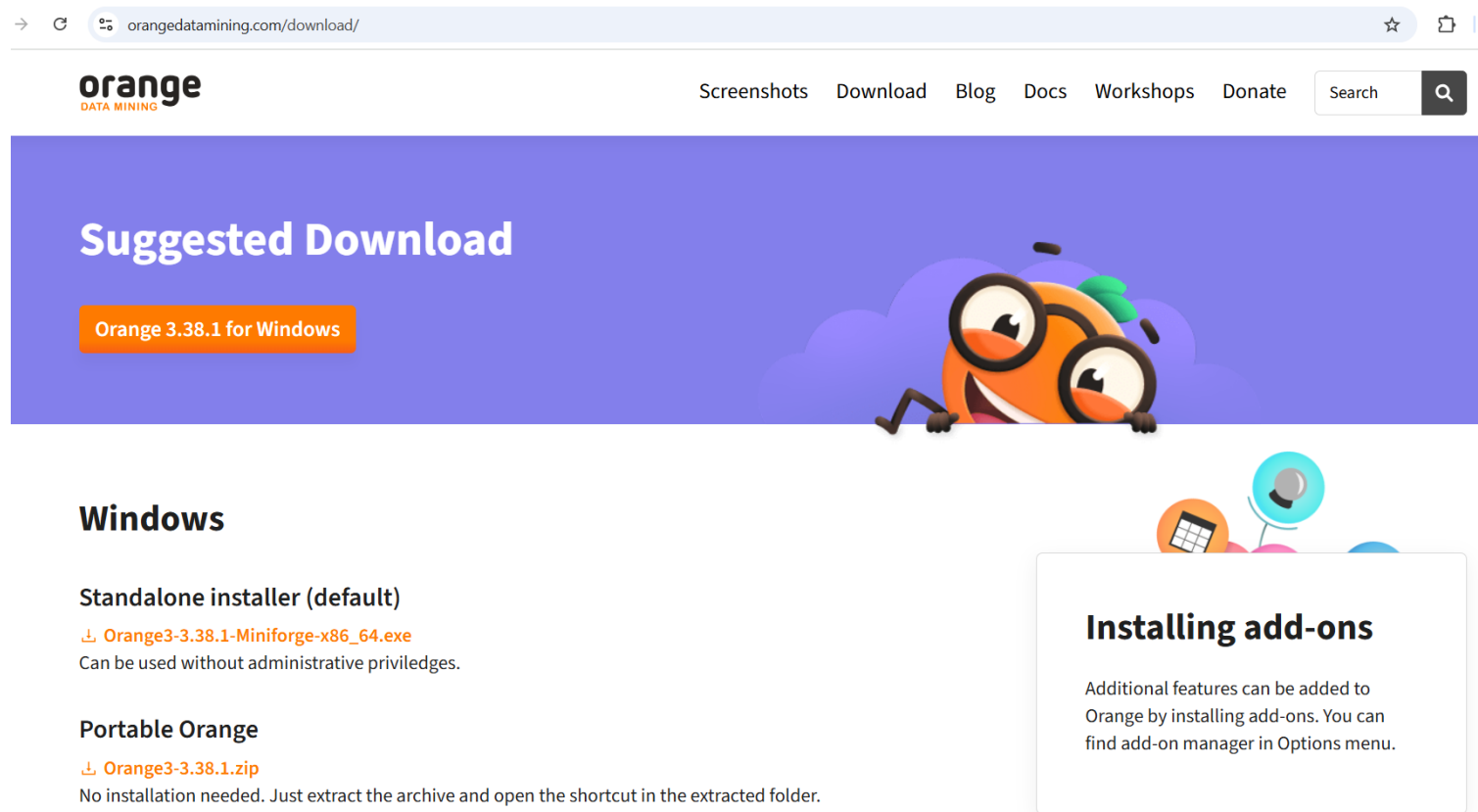
# Orange

- Orange is a powerful platform to perform data analysis and visualization, see data flow and become more productive. It provides a clean, open-source platform and the possibility to add further functionality for all fields of science.



# Orange, Download and Install

- <https://orangedatamining.com/download/>



The screenshot shows the Orange Data Mining website's download page. The browser address bar displays "orangedatamining.com/download/". The website header includes the "orange DATA MINING" logo and navigation links: "Screenshots", "Download", "Blog", "Docs", "Workshops", and "Donate". A search bar is located on the right. The main content area features a large purple banner with the text "Suggested Download" and a button labeled "Orange 3.38.1 for Windows". Below this, the "Windows" section is divided into two options: "Standalone installer (default)" with a download link for "Orange3-3.38.1-Miniforge-x86\_64.exe" and a note that it can be used without administrative privileges, and "Portable Orange" with a download link for "Orange3-3.38.1.zip" and a note that no installation is needed. A side panel titled "Installing add-ons" explains that additional features can be added by installing add-ons and that an add-on manager is available in the Options menu. The Orange mascot is visible on the right side of the page.

orange  
DATA MINING

Screenshots Download Blog Docs Workshops Donate Search

## Suggested Download

Orange 3.38.1 for Windows

## Windows

### Standalone installer (default)

📄 [Orange3-3.38.1-Miniforge-x86\\_64.exe](#)  
Can be used without administrative privileges.

### Portable Orange

📄 [Orange3-3.38.1.zip](#)  
No installation needed. Just extract the archive and open the shortcut in the extracted folder.

### Installing add-ons

Additional features can be added to Orange by installing add-ons. You can find add-on manager in Options menu.

# Rapidm in er

---



## Depth for Data Scientists, Simplified for Everyone Else

Join the 40,000+ global organizations in every industry who use the RapidMiner data science platform to drive revenue, reduce costs, and avoid risk.

- RapidMiner provides data mining and machine learning procedures including: data loading and transformation (ETL), data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment.
- RapidMiner provides a GUI to design and execute analytical workflows.





# RAPIDMINER

---

## The Hard Way

- › Slow to collaborate & transform
- › Acquire highly specialized expertise
- › Choose complexity or oversimplification
- › Opaque, fragmented processes
- › Commit to a specific ecosystem

## The RapidMiner Way

- › Transformational business impact
- › Upskill your organization
- › Depth for data scientists, simplified for everyone else
- › Easy to trust, tune & explain
- › Future-proof innovation, portability & extensibility





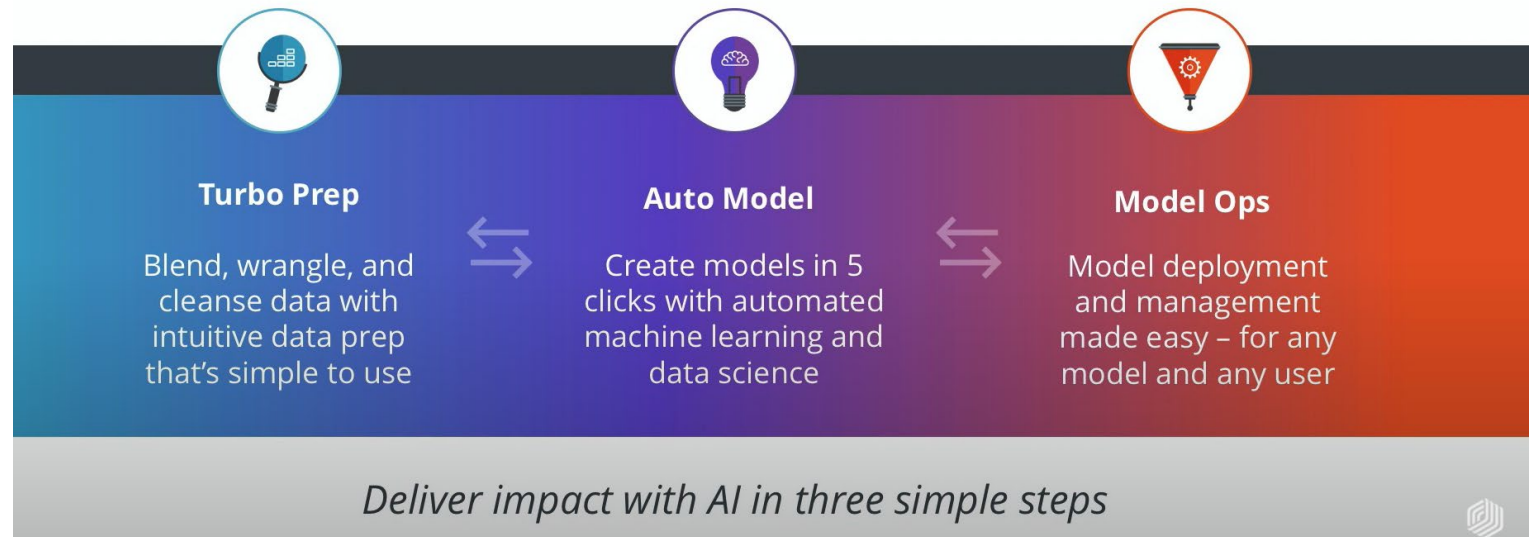
# RAPID MINER

---

[WWW.RAPIDMINER.COM](http://WWW.RAPIDMINER.COM),

[WWW.YOUTUBE.COM](http://WWW.YOUTUBE.COM) RAPID MINER

## RapidMiner Offers a Complete Path to Fully Automated Data Science



# RAPID MINER



• ~~HELD~~CK



**FORRESTER**

RapidMiner is a Leader in The Forrester Wave: Multimodal Predictive Analytics & Machine Learning Solutions, Q3 2020

**Gartner**

RapidMiner is a Visionary in the 2021 Gartner Magic Quadrant for Data Science and Machine Learning Platforms



RapidMiner is a June 2020 Gartner Peer Insights Customers' Choice for Data Science and Machine Learning Platforms for the third time in a row

**G<sup>2</sup> | CROWD**

RapidMiner is the Highest Rated, Easiest to Use Data Science and Machine Learning Platform and was named a Leader in G2's Spring 2021 Report.



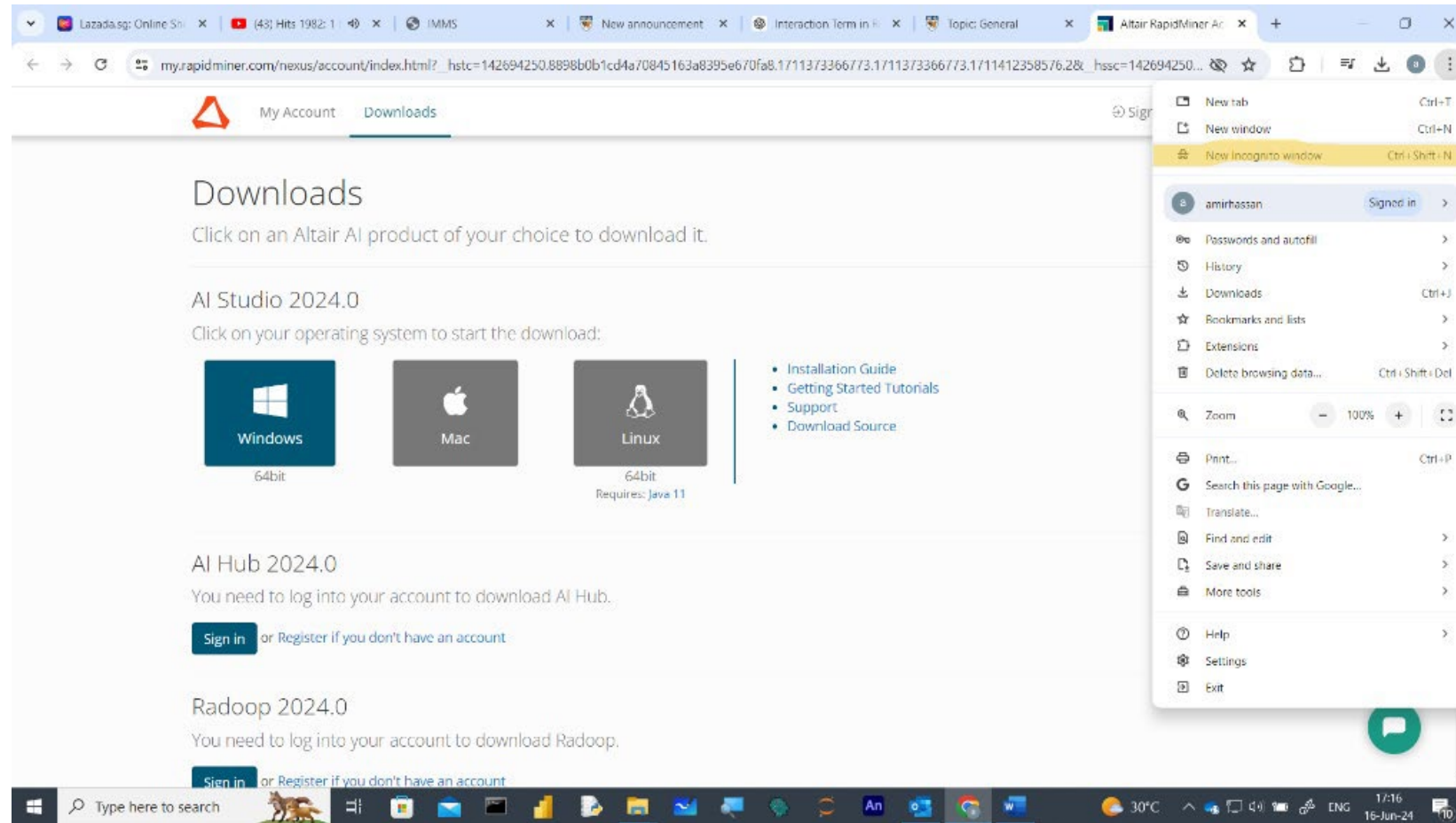
Rapidminer

# Download and Install

---

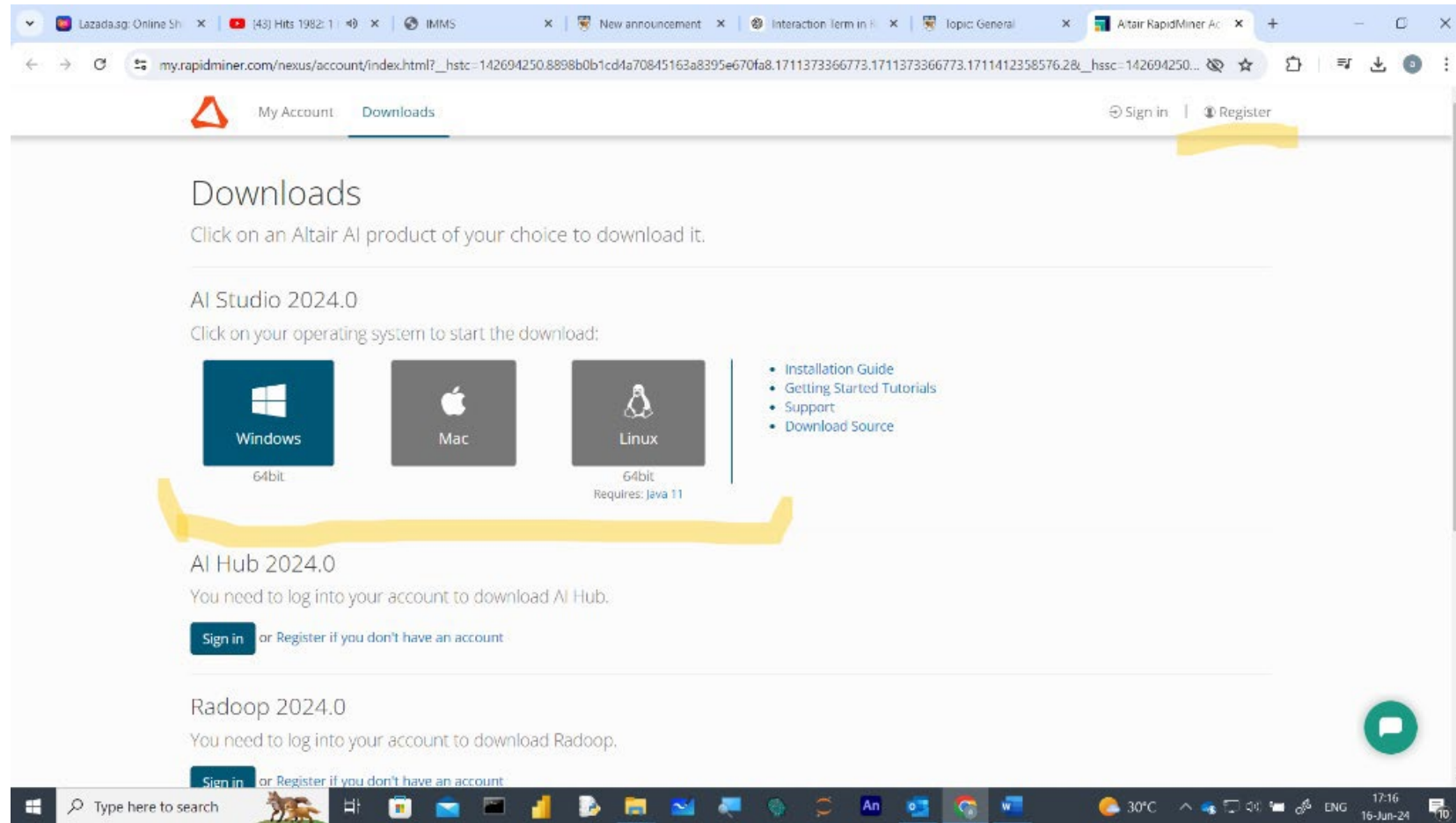
- To download Rapidminer:
- [https://my.rapidminer.com/nexus/account/index.html?\\_hstc=142694250.8898b0b1cd4a70845163a8395e670fa8.1711373366773.1711373366773.1711412358576.2&\\_hssc=142694250.1.1711412358576&\\_hsfp=676704640](https://my.rapidminer.com/nexus/account/index.html?_hstc=142694250.8898b0b1cd4a70845163a8395e670fa8.1711373366773.1711373366773.1711412358576.2&_hssc=142694250.1.1711412358576&_hsfp=676704640)
- You need to have a Rapidminer account.
- Try this website to create a Rapidminer account instead of the main webpage:
- [https://my.rapidminer.com/nexus/account/index.html?\\_hstc=142694250.3188faa95341077fb0a799ae44e201f4.1714093534218.1714093534218.1714096391239.2&\\_hssc=142694250.4.1714096391239&\\_hsfp=2414989234#signup](https://my.rapidminer.com/nexus/account/index.html?_hstc=142694250.3188faa95341077fb0a799ae44e201f4.1714093534218.1714093534218.1714096391239.2&_hssc=142694250.4.1714096391239&_hsfp=2414989234#signup)
- Tick "For Educational Purpose"
- Using incognito sometimes is better if your browser is Chrome.

# Download and Install



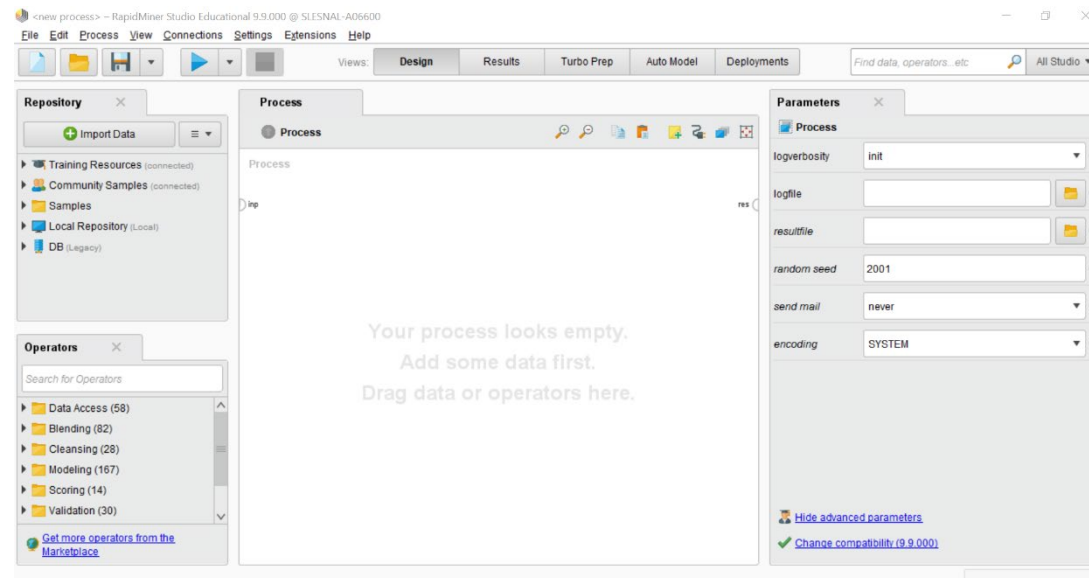


# Download and Install



# HOW TO START?

- Mind your OS
- Install it
  - The basic **RapidMiner Studio** is free.
- Run it !!!
- [www.rapidminer.com](http://www.rapidminer.com)



# HOW TO START?

The image shows the 'Start' tab of the RapidMiner Studio interface. It is titled 'Welcome to RapidMiner Studio!' and has three sub-tabs: 'Start', 'Recent', and 'Learn'. The 'Start' tab is active and displays two main sections: 'Start with' and 'Choose a template to start from'.

**Start with**

- Blank Process**: Start a new process from scratch in the design view.
- Turbo Prep**: Prepare your data interactively: transform, clean and combine data sets.
- Auto Model**: Build and optimize models using automated machine learning.

**Choose a template to start from**

- Churn Modeling**: Predict which of your customers will churn and why with a decision tree.
- Direct Marketing**: Predict response to campaigns and increase the conversion rate of your campaign.
- Credit Risk Modeling**: Model credit default risk by training an optimized Support Vector Machine (SVM) model.
- Market Basket Analysis**: Find products frequently purchased together and turn them into rules for recommendations.
- Predictive Maintenance**: Model equipment failures to schedule maintenance pre-emptively.
- Price Risk Clustering**: Cluster price developments using X-Means to unveil price-risk-relationships.
- Lift Chart**: Create a lift chart to visualize the improvement that a model provides compared to guessing.
- Operationalization**: Embed predictive models into business processes to trigger the right actions automatically.
- Outlier Detection**: Detect anomalies in data resulting from a chemical analysis of wines.
- Geographic Distances**: Calculate the nearest antenna to a given client position.
- Medical Fraud Detection**: Detect medical fraud based on patient information.
- Web Analytics**: Score users as high/low value based on their page visits on a web site.

**Callouts:**

- Recent Actions**: Points to the 'Recent' sub-tab.
- Start with this**: Points to the 'Start with' section.
- Advanced mode**: Points to the 'Auto Model' option.
- Ready to use ML and DA Templates. We'll see these later**: Points to the 'Choose a template to start from' section.



Views:

Design

Results

Turbo Prep

Auto Model

Deployments

Find data, operators...etc



## Repository

+ Import Data

- ▶ Training Resources (connected)
- ▶ Community Samples (connected)
- ▶ Samples
- ▶ Local Repository (Local)
- ▶ DB (legacy)

## Operators

Search for Operators

- ▶ Data Access (58)
- ▶ Blending (82)
- ▶ Cleansing (28)
- ▶ Modeling (167)
- ▶ Scoring (14)
- ▶ Validation (30)

[Get more operators from the Marketplace](#)

## Process

Process

inp

res

Your process looks empty.  
Add some data first.  
Drag data or operators here.

RAPIDMINER EDUCATIONAL

## Parameters

Process

logverbosity init

logfile

resultfile

random seed 2001

send mail never

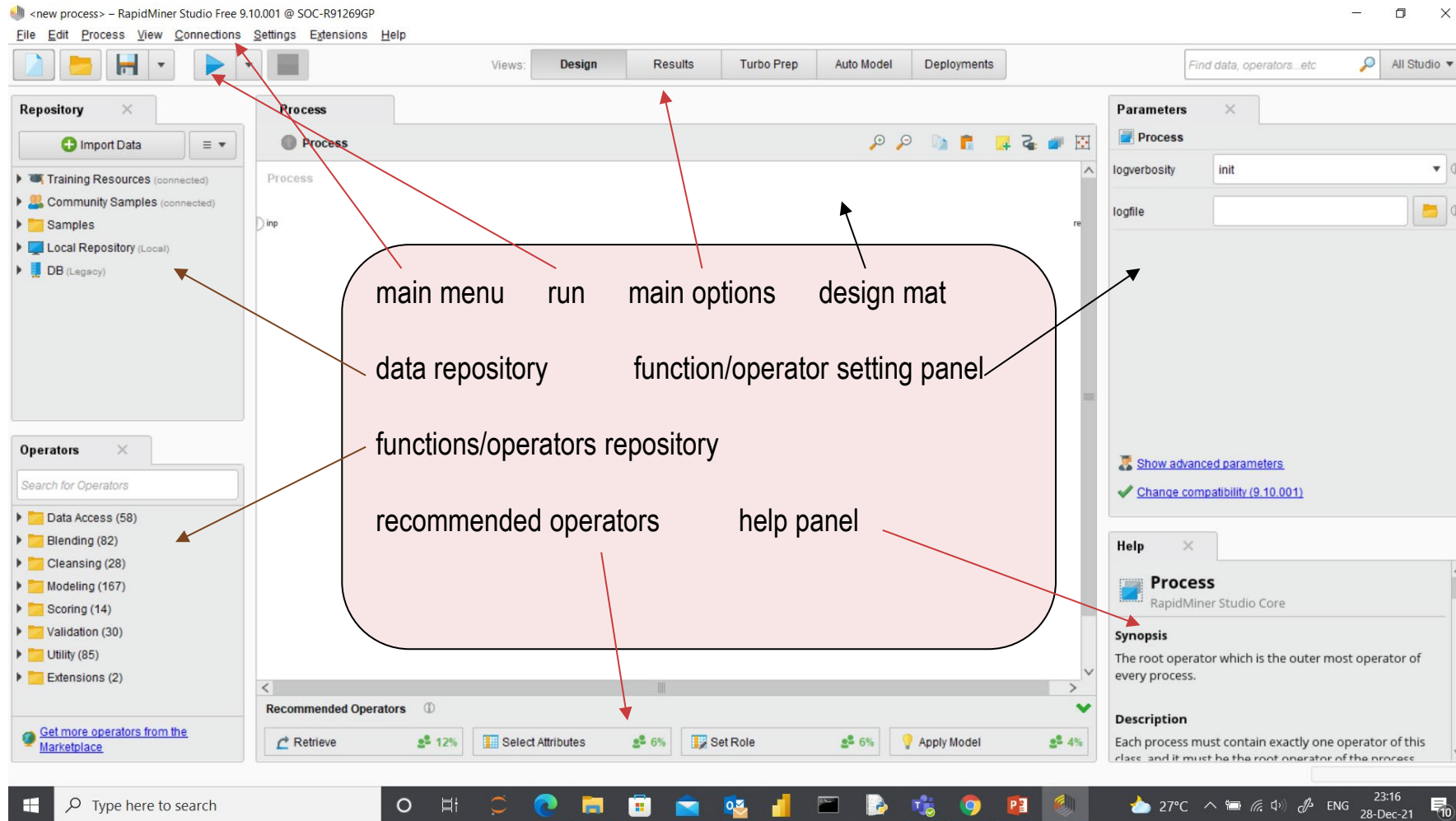
encoding SYSTEM

[Hide advanced parameters](#)

[Change compatibility \(9.9.000\)](#)



# RAPIDMINER MAIN PAGE



# Rapidminer Main page

The screenshot shows the Rapidminer Studio interface with the following components and annotations:

- Repository:** Contains data sources like Training Resources, Community Samples, Samples, Local Repository, and DB. Annotation: "New design load save a model" points to the Import Data button and the Repository tree.
- Process:** The central workspace for building a process. Annotation: "Design a model and see the results Run it" points to the Design and Results tabs.
- Design Canvas:** The central area where operators are added. Annotation: "Click on an operator on your design canvas and set its parameters here" points to the Parameters panel.
- Operators:** A list of available operators categorized by function (Data Access, Blending, Cleansing, Modeling, Scoring, Validation, Utility, Extensions). Annotation: "A good way to find an operator is typing a part of its name here" points to the search bar.
- Parameters:** A panel for configuring the selected operator. It shows parameters like logverbosity and logfile.
- Help:** A panel providing information about the selected operator, including a synopsis and description.

**DESIGN CANVAS**

# Libraries

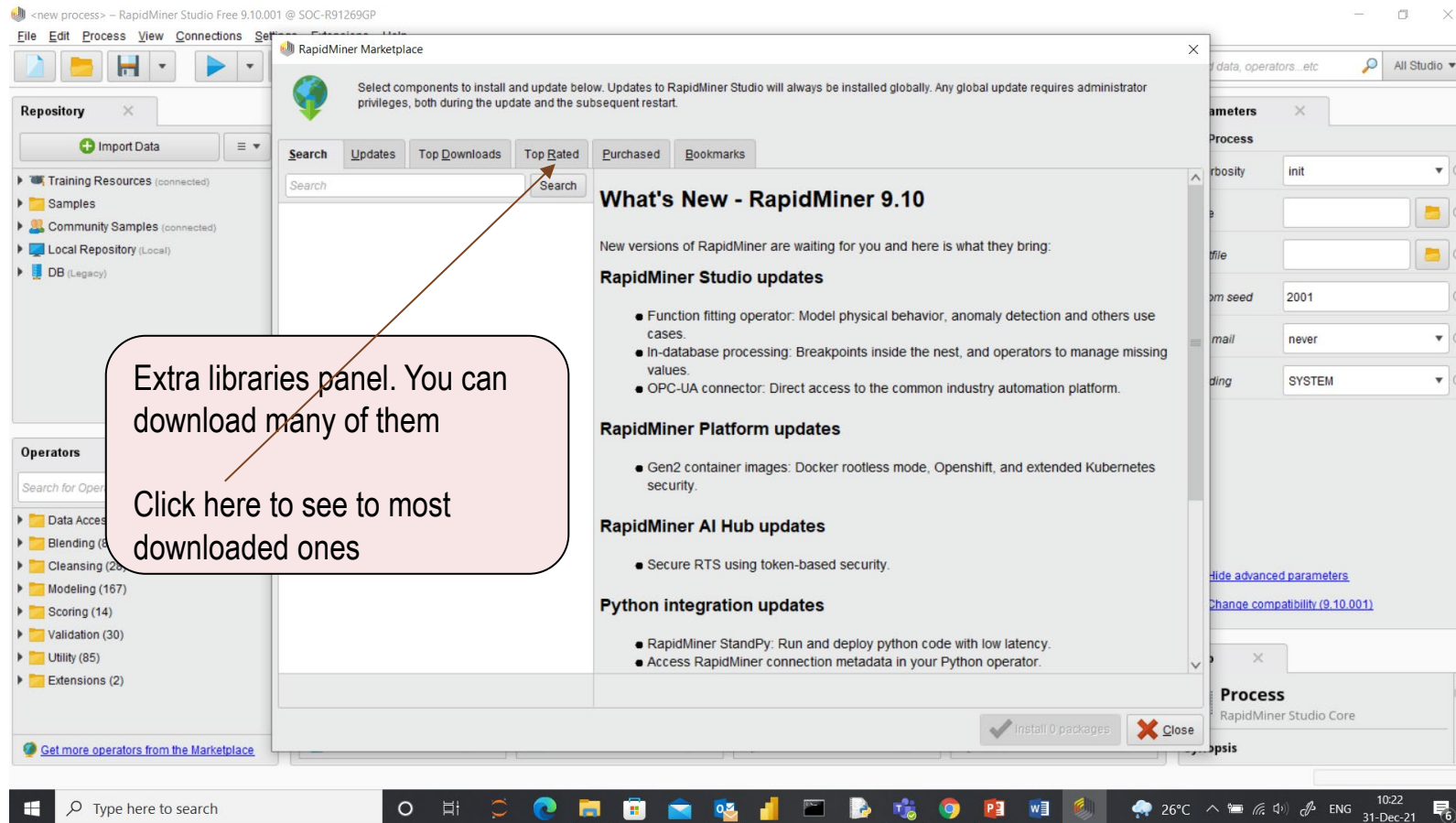
The screenshot displays the RapidMiner Studio Free 9.10.001 interface. The main window is titled "<new process> - RapidMiner Studio Free 9.10.001 @ SOC-R91269GP". The menu bar includes File, Edit, Process, View, Connections, Settings, Extensions, and Help. The toolbar contains icons for file operations and execution. The interface is divided into several panels:

- Repository:** Located on the left, it shows a tree view of data sources: Training Resources (connected), Community Samples (connected), Samples, Local Repository (Local), and DB (Legacy). It includes an "Import Data" button.
- Operators:** Also on the left, it features a search bar "Search for Operators" and a list of operator categories with counts: Data Access (58), Blending (82), Cleansing (28), Modeling (167), Scoring (14), Validation (30), Utility (85), and Extensions (2). A link "Get more operators from the Marketplace" is at the bottom.
- Process:** The central workspace, currently empty, with a "Process" tab selected.
- Parameters:** On the right, it shows settings for the "Process" operator, including "logverbosity" set to "init" and a "logfile" field. It also has links for "Show advanced parameters" and "Change compatibility (9.10.001)".
- Help:** At the bottom right, it provides a "Synopsis" and "Description" for the "Process" operator.

A pink callout box with a black border is positioned over the Operators panel. It contains the text "Rapidminer v.9 standard library" and "Click here for extra libraries". Two arrows point from this box to the "Data Access (58)" and "Extensions (2)" categories in the Operators list.

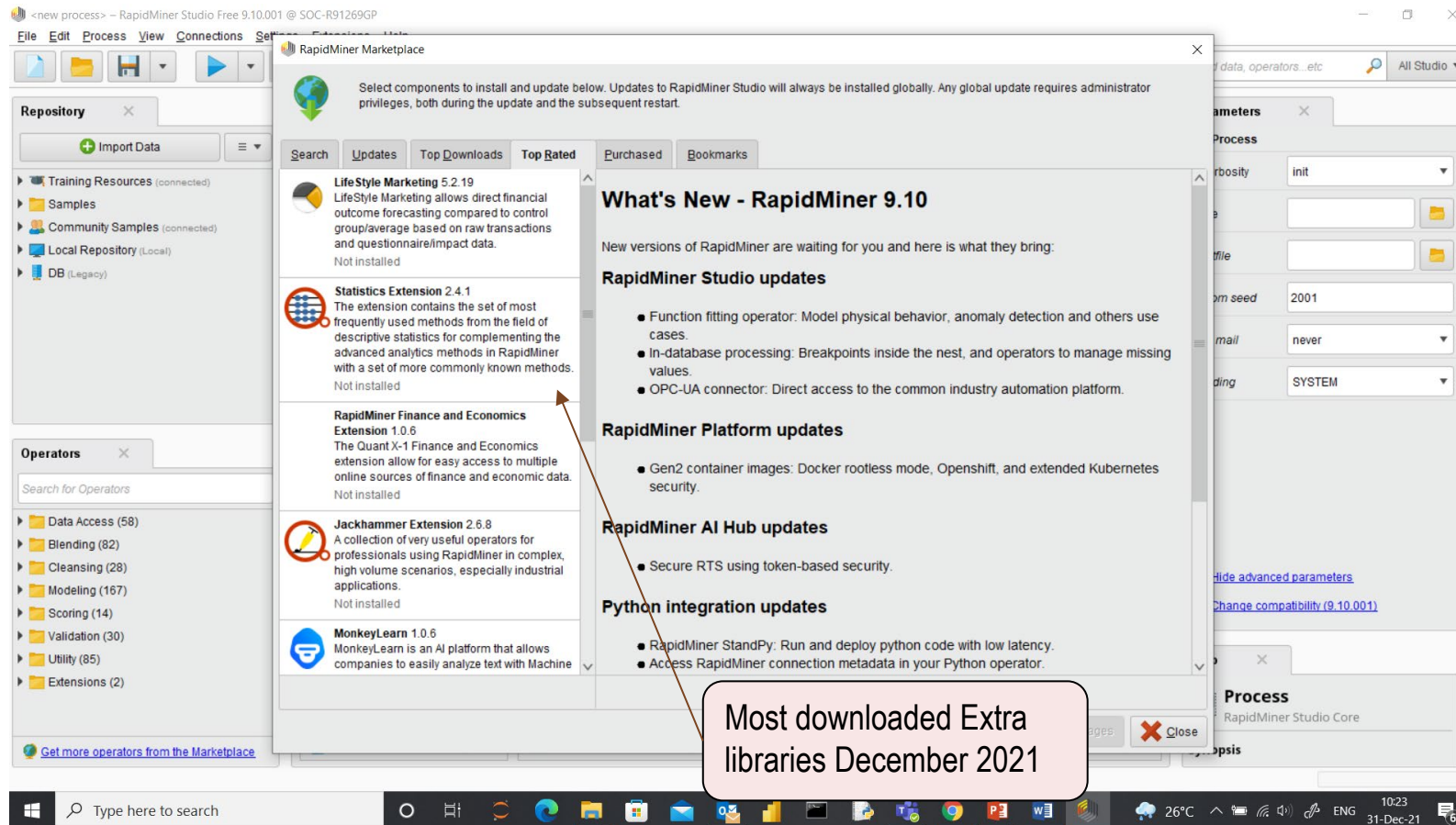
At the bottom of the screen, the Windows taskbar is visible, showing the search bar "Type here to search" and various application icons. The system tray on the right indicates a temperature of 27°C, time of 23:16, date of 28-Dec-21, and language set to ENG.

# Libraries





# Libraries



The background is a blurred image of a document. It features a line graph with several data points connected by lines. A pen is visible in the upper right corner, resting on the paper. The overall color palette is light blue and white.

# Data Preprocessing

---

# Data Preprocessing Example

---

- Data file: celest-bad-25.csv
- Python notebook preprocessing-25-1.ipynb
- Goals
  - Handling missing values by removing
  - Handling missing values by replacement
  - Data normalization
  - Outlier handling

# Data file

name	height	weight	gender	DOB
Alicia Lina	178	115f		Jun-81
Alicia Keys	164	120f		Jan-81
Angela Jolie	127	122f		Jun-75
Ama Kunkwa	127	120f		Jun-81
Ashley Judd	102	125f		Apr-68
Anil Lavigne	155	113f		Sep-84
Bar Refaeli	174	128f		Jun-85
Byrne Kowles	166	129f		Sep-81
Elle Lively	127	f		Aug-87
Gemma Ffrench	160	113f		Apr-72





# The End...

Tutorial 1