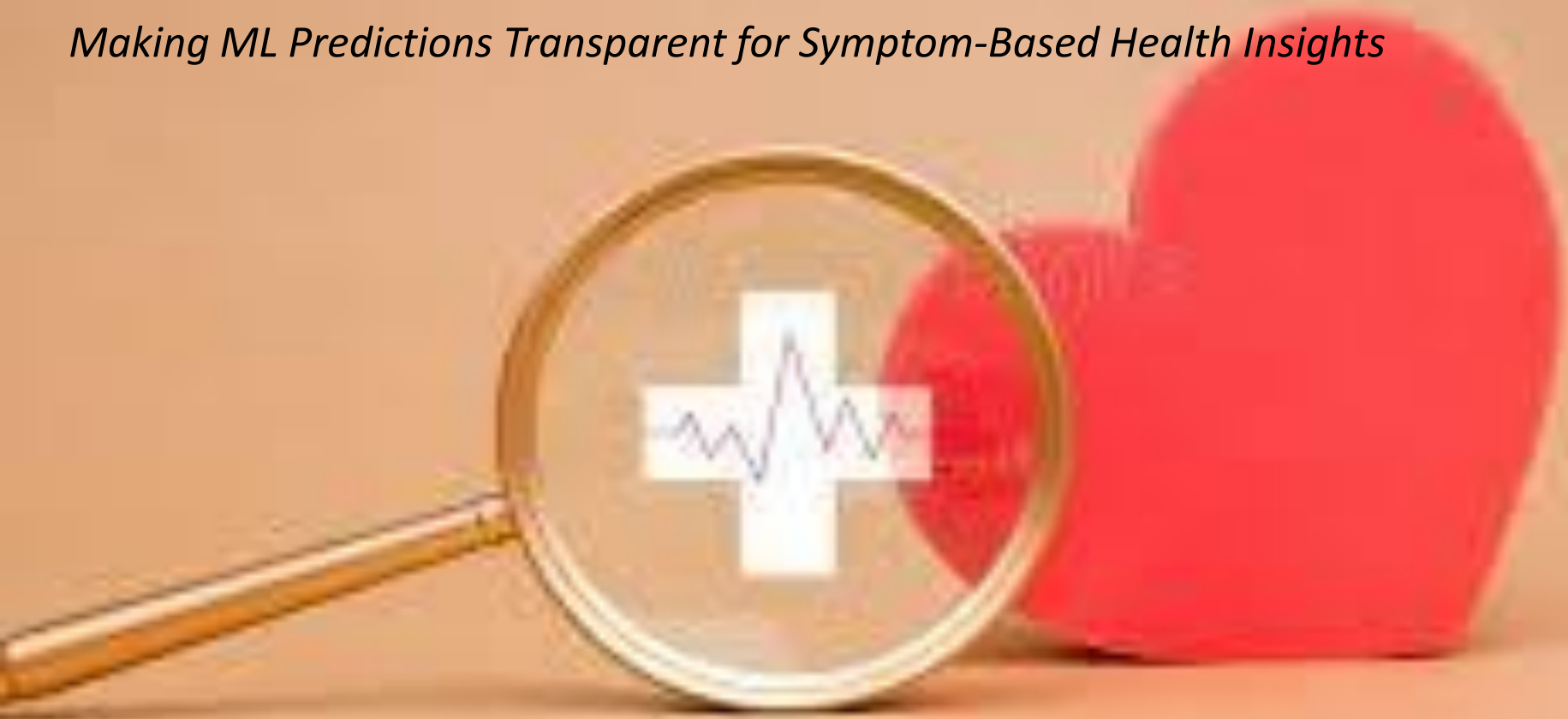


Demystifying Disease Risk: A Visual predictor

Making ML Predictions Transparent for Symptom-Based Health Insights



The “Black Box” Problem in Health AI



The Power of AI in Healthcare

- Machine Learning (ML) is **transforming** healthcare.
- It's **used for everything** from predicting disease outbreaks to aiding in diagnoses and personalizing treatment plans.

The Challenge

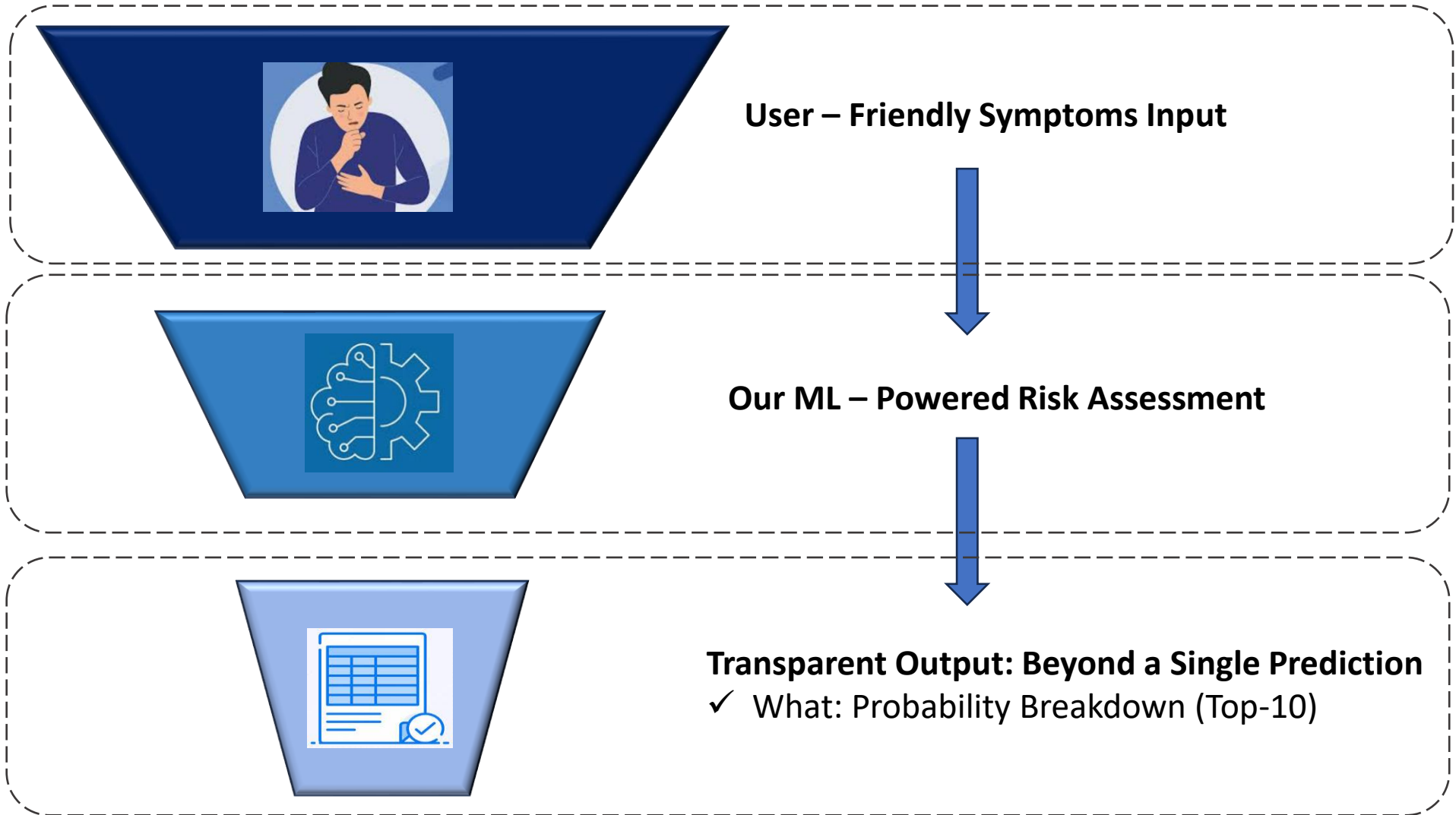
- Many powerful ML models are **complex** inherently opaque.
- It's **hard to understand** *how* they arrive at their predictions
- They **act like a “Black box”** – inputs go in, predictions come out, but the internal logic is hidden.

Impact in Healthcare


- Leads to distrust and misunderstanding.
- Limits effective action based on AI insights.

OUR SOLUTION: Apply Data Visualization & Analytics to make AI transparent

Our Solution: The Visual Predictor (High-Level)



What We Delivered

 **Disease-Symptom Risk Predictor**

Disclaimer: This tool is for informational purposes only and should NOT be considered a substitute for professional medical advice, diagnosis, or treatment. Always seek the advice of your physician or other qualified health provider with any questions you may have regarding a medical condition. Do not disregard professional medical advice or delay in seeking it because of something you have read on this tool.

This application was developed as a project for a Data Analytics Bootcamp and is intended for demonstration purposes only.

Tick the symptoms you observe, then click Predict.

<input type="checkbox"/> Abdominal pain	<input type="checkbox"/> Abnormal menstruation	<input type="checkbox"/> Acidity	<input type="checkbox"/> Acute liver failure
<input type="checkbox"/> Altered sensorium	<input type="checkbox"/> Anxiety	<input type="checkbox"/> Back pain	<input type="checkbox"/> Belly pain
<input type="checkbox"/> Blackheads	<input type="checkbox"/> Bladder discomfort	<input type="checkbox"/> Blister	<input type="checkbox"/> Blood in sputum
<input type="checkbox"/> Bloody stool	<input type="checkbox"/> Blurred and distorted vision	<input type="checkbox"/> Breathlessness	<input type="checkbox"/> Brittle nails
<input type="checkbox"/> Bruising	<input type="checkbox"/> Burning micturition	<input type="checkbox"/> Chest pain	<input type="checkbox"/> Chills
<input type="checkbox"/> Cold hands and feet	<input type="checkbox"/> Coma	<input checked="" type="checkbox"/> Congestion	<input type="checkbox"/> Constipation

Predict

Most likely diagnosis: GERD

Full probability breakdown (top-10)

Disease	Probability
GERD	14.93%
Hypertension	6.48%
Heart attack	6.33%
Bronchial Asthma	4.66%
Urinary tract infection	3.39%
Allergy	3.37%
Acne	3.19%
Gastroenteritis	3.13%
AIDS	3.07%
Fungal infection	3.05%

Powered by scikit-learn · Data courtesy of kaushil268 / Kaggle

- **Intuitive Symptom Input:** User-friendly web interface with symptom selection
- **ML Model for Risk Prediction:** Identifies potential disease based on selected symptoms
- **Core Transparency Features**
 - Full Probability Breakdown (Top 10): Shows model's confidence across top diseases

Core Technologies & Libraries

Data & ML Core



Web Delivery & UI Interface



Visuals & Collaboration



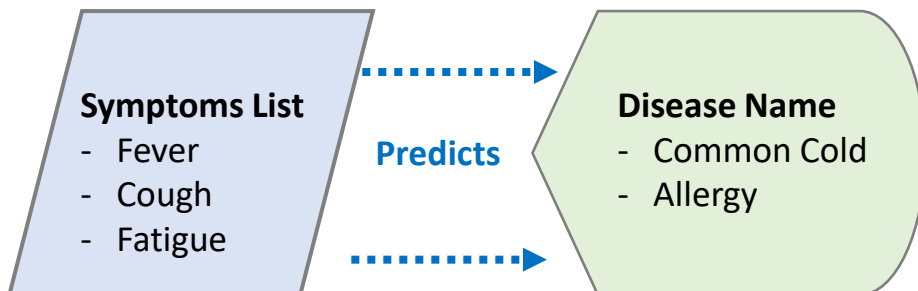
Our Data: Symptoms to Disease

kaggle

- ✓ **Data Source:** Kaggle's Disease-Symptom Description Dataset (*Clear symptom-to-disease mappings needed for our predictive model*)

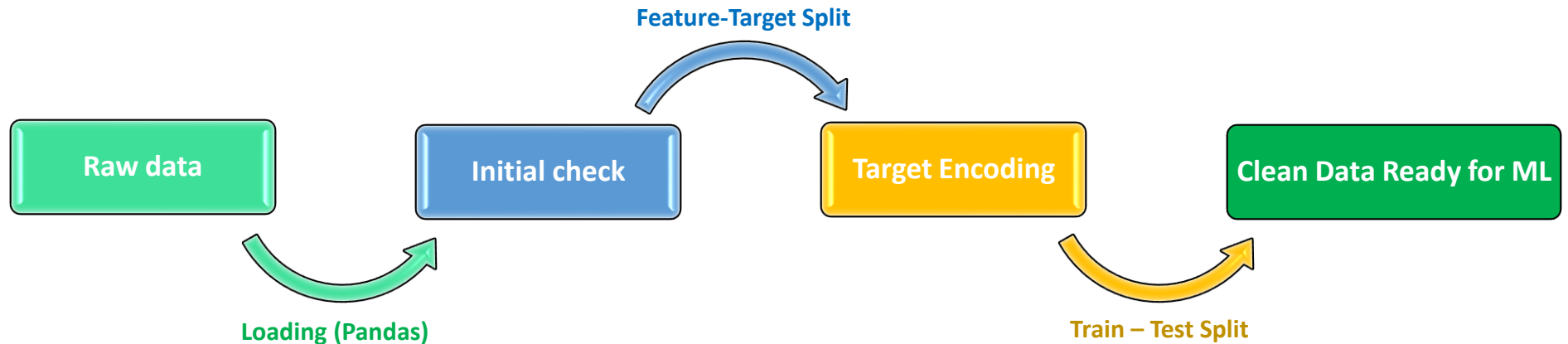
Δ Disease	Δ Symptom_1	Δ Symptom_2	Δ Symptom_3	Δ Symptom_4
Diseases that may be present	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease
41 unique values	vomiting 17% fatigue 14% Other (3408) 69%	vomiting 18% fatigue 8% Other (3648) 74%	fatigue 15% high_fever 7% Other (3870) 79%	high_fever 8% [null] 7% Other (4194) 85%
Fungal_infection	itching	skin_rash	nodal_skin_eruptions	
Fungal_infection	itching	skin_rash	nodal_skin_eruptions	dischromic_patches
Allergy	continuous_sneezing	shivering	chills	watering_from_eyes
Allergy	shivering	chills	watering_from_eyes	
Allergy	continuous_sneezing	chills	watering_from_eyes	
Allergy	continuous_sneezing	shivering	watering_from_eyes	
Allergy	continuous_sneezing	shivering	chills	
Allerov	shivering	chills	watering from eyes	

- ✓ **Structure:** Symptoms directly mapped to disease prognoses. (*Initially CSVs, pre-processed into an SQLite database for efficient access. Allowing us to directly use symptom presence as features*)



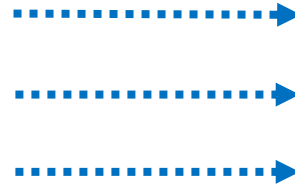
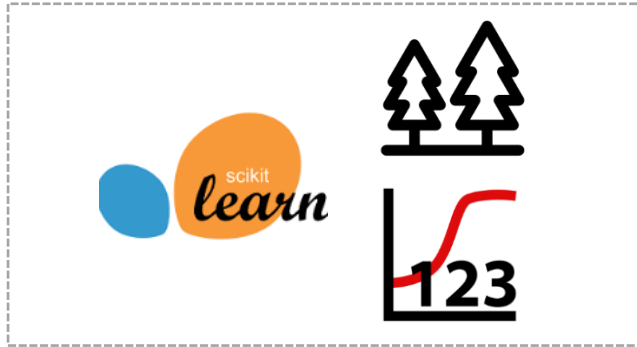
- ✓ **Relevance:**
Ideal for our symptom-based disease prediction model
- ✓ **Format:**
Pre-cleaned and stored in an SQLite database

Preparing Data for Prediction



- **Loading & Initial Checks:** From SQLite database into Pandas DataFrame.
- **Feature – Target Split:** Separated symptoms (features) from disease (target)
- **Target Encoding:** Converted text disease names to numerical labels (and back for output)
- **Train-test Split:** Divided data for fair model evaluation and to prevent overfitting.

Choosing Our Prediction Engine



Choose Model

Problem Type: Supervised Classification.

Algorithms Explored (Scikit-learn):

- ✓ **Logistic Regression:** Efficient, probabilistic.
- ✓ **Random Forest:** Robust, high accuracy.

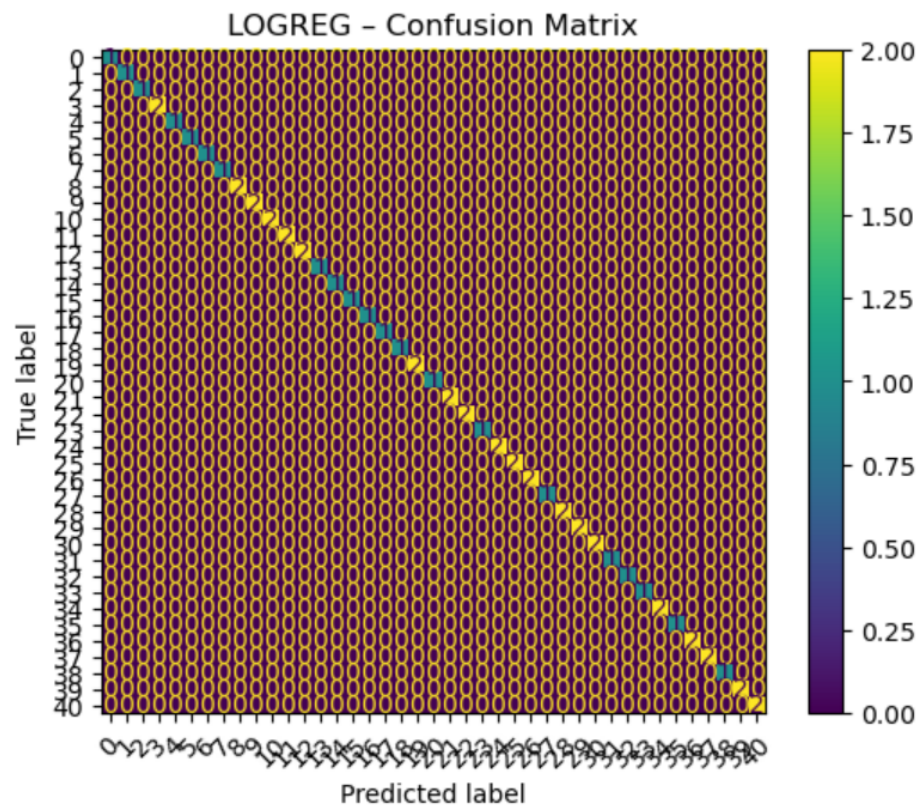
Selection

Chose the best performer based on predictive accuracy and suitability for probability output.



Training: Model trained on our prepared data.

Evaluating Our Model's Effectiveness



➤ **Assessing Reliability:** Crucial step to ensure trustworthy predictions.

➤ **Key Metrics:**

➤ **Accuracy:**

logreg	accuracy: 1.000
rf	accuracy: 1.000

➤ **Confusion Matrix:** Provides detailed breakdown of correct vs. incorrect classifications.

OVERALL: Our model demonstrates strong predictive performance.

Our Virtual Predictor in Action!

Live Demo

Future Enhancements



- Enhanced Symptom Input (e.g., search, NLP)



- More Advanced Visual Explanations



- User Feedback & Model Refinement Loop



- Public Deployment

Overcoming Hurdles & Key Takeaways

Challenges Faced:

- Managing scope in a 2 week-sprint
- Integrating ML backend with Streamlit frontend
- Designing effective demystification visuals

Key Learnings:

- Practical application of the full ML workflow
- Importance of user-centric AI design for transparency.



A vibrant, 3D-style graphic of the words "THANK YOU!" in a playful, bold font. Each letter is a different color: 'T' is blue, 'H' is purple, 'A' is pink, 'N' is orange, 'K' is yellow, 'Y' is green, and 'O' is teal. The exclamation mark is also teal. The text is surrounded by a cloud of small, multi-colored dots in shades of blue, yellow, orange, and purple, giving it a festive, confetti-like appearance.

THANK YOU!