NutriScanner

User Manual

AI-Powered Nutritional Information Extractor

Extract allergens and nutritional values from food product PDFs

Key Features

PDF Support	Multi-language	AI-Powered
10 Allergens	Nutritional Data	Real-time

Version 1.0 — October 26, 2025

Contents

1 Introduction

1.1 Welcome to NutriScanner

NutriScanner is an advanced web application that uses artificial intelligence to automatically extract allergen information and nutritional values from food product PDF documents. Whether you're dealing with product labels in Hungarian, English, or both languages, NutriScanner makes it easy to get the information you need quickly and accurately.

Information

NutriScanner supports both digital PDFs and scanned images, handling everything from perfectly formatted nutrition labels to handwritten or poorly scanned documents.

1.2 What NutriScanner Can Do

- **Allergen Detection:** Automatically identifies all 10 EU-regulated allergens.
- **Nutritional Analysis:** Extracts energy, fat, carbohydrates, sugar, protein, and sodium values.
- **Bilingual Support:** Works seamlessly with Hungarian and English documents.
- **AI-Powered:** Choose between different backend AI engines (configurable).
- **Real-Time Progress:** Watch extraction happen live with progress updates.
- **Confidence Scoring:** Know how reliable the extracted data is.

1.3 Who Should Use NutriScanner

NutriScanner is perfect for:

- Food manufacturers
- Nutritionists and dietitians
- Quality control teams
- Restaurant owners
- Food importers/exporters
- Health-conscious consumers
- Allergy sufferers
- Researchers

2 Getting Started

2.1 System Requirements

To use NutriScanner, you need:

- **Web Browser:** Modern browser (Chrome, Firefox, Safari, Edge)
- **Internet Connection:** Required for AI processing
- **PDF Files:** Maximum 5MB in size (recommended)
- **Supported Formats:** Digital PDFs or scanned images saved as PDF

Tip

For best results, use PDFs with clear, readable text. Scanned images should be at least 200 DPI resolution.

2.2 Accessing NutriScanner

1. Open your web browser.
2. Navigate to the NutriScanner website.
3. The application loads without installation; sign-in depends on your deployment.

2.3 Understanding the Interface

When you first open NutriScanner, you'll see:

Model Selector Choose the AI backend to use for extraction.

Upload Area Drag and drop your PDF or click to browse files.

Extract Button Start the AI analysis process.

Documentation Access help and additional information.

3 How to Use NutriScanner

3.1 Complete Step-by-Step Guide

Step 1: Select Your AI Model **Choose the AI engine that will analyze your PDF:**

Model	Best For
OpenAI GPT-4 style model	Critical data, maximum accuracy, complex documents
Google Gemini-style model	Fast processing, cost-effective, clear documents

Recommendation: Start with the highest-accuracy model for your first extraction.

Step 2: Upload Your PDF **Two ways to upload:**
1. **Drag & Drop:** Drag your PDF file directly onto the upload area.
2. **Click to Browse:** Click anywhere in the upload area to open a file picker.

File Requirements:
- Format: PDF only (.pdf)
- Size: Maximum 5MB
- Content: Food product labels with allergen or nutritional information

Step 3: Extract Information **Start the AI analysis:**
1. Click the **Extract Information** button.
2. Watch the real-time progress bar (0–100%).
3. See status updates as data is extracted.
4. The application notifies when extraction completes.

Typical extraction time: 10–30 seconds depending on document complexity.

Step 4: Review Your Results **Your extracted data will display in organized sections:**
- **Product Information:** Name, language detected, confidence level.
- **Allergen Table:** Visual indicators (check = present, cross = absent).
- **Nutritional Values:** All extracted nutrients with original units.

Step 5: Scan Another Document **Process additional PDFs:**
- Click the **Scan Another** button to reset to the upload screen.
- Previously selected model may be remembered depending on settings.

4 Understanding Your Results

4.1 Allergen Information

NutriScanner detects all **10 EU-regulated allergens**:

Symbol	English	Hungarian
✓	Gluten	Glutén
✓	Egg	Tojás
✓	Crustaceans	Rák
✓	Fish	Hal
✓	Peanut	Földimogyoró
✓	Soy	Szója
✓	Milk	Tej
✓	Tree Nuts	Diófélék
✓	Celery	Zeller
✓	Mustard	Mustár

Information

Present (check): Document explicitly states "contains" or lists the allergen in ingredients.
Absent (cross): Allergen is not mentioned or only shown as "may contain traces".
Note: Cross-contamination warnings do *not* trigger a "present" status.

4.2 Nutritional Values

NutriScanner extracts these nutritional components:

Energy Displayed in kJ and/or kcal (e.g., "1500 kJ / 350 kcal").

Fat Grams of total fat per 100g / 100ml.

Carbohydrates Total carbohydrate content in grams.

Sugar "Of which sugars" value in grams.

Protein Protein content in grams.

Sodium Sodium/salt content in mg or grams.

Tip

Units are preserved exactly as they appear in the document. NutriScanner does not automatically convert units.

4.3 Confidence Levels

Every extraction includes a confidence assessment:

Level	Meaning
High	Clear, readable document with complete information.
Medium	Mostly readable, some interpretation needed.
Low	Poor quality, degraded, or incomplete data — verify manually.

Warning

Always verify Low confidence results manually. Consider re-scanning the document or using a different model.

4.4 Language Detection

NutriScanner automatically detects the document language:

- **Hungarian:** Document uses Hungarian terminology.
- **English:** Document uses English terminology.
- **Both:** Bilingual label with both languages.
- **Unknown:** Blank, illegible, or other language.

5 Tips & Best Practices

5.1 Getting the Best Results

Tip

1. Use high-quality scans: 200+ DPI for scanned images.
2. Ensure good lighting: Avoid shadows and glare in photos.
3. Full page visibility: Include the entire nutrition label.
4. Straight alignment: Keep documents as flat and straight as possible.
5. Clear text: Avoid blurry or out-of-focus images.

5.2 Choosing the Right AI Model

Scenario	Recommended Model	Reason
Critical allergen info	Highest-accuracy model	Best for complex documents
Batch processing	Faster model	Cost-effective for clear docs
Poor quality scans	Highest-accuracy model	Better OCR capabilities
Clear, digital PDFs	Faster model	Equally accurate for simple layouts
Complex layouts	Highest-accuracy model	Better structure understanding

5.3 Handling Common Document Types

5.3.1 Professional Nutrition Labels

- Expected result: High confidence.
- Model choice: Either high-accuracy or fast model.
- Extraction time: 10–15 seconds.

5.3.2 Scanned Product Packages

- Expected result: Medium to High confidence.
- Model choice: High-accuracy recommended.
- Extraction time: 15–25 seconds.

5.3.3 Handwritten or Degraded Labels

- Expected result: Low to Medium confidence.
- Model choice: High-accuracy strongly recommended.
- Extraction time: 20–30 seconds.
- Note: Verify results manually.

6 Troubleshooting

6.1 Common Issues and Solutions

6.1.1 Upload Errors

Warning

Error: Only PDF files under 5MB are allowed
Cause: File is too large or wrong format.
Solutions:
- Compress your PDF using PDF software.
- Split multi-page PDFs into separate files.
- Convert images to PDF format first.
- Reduce image resolution in scanned PDFs.

6.1.2 Extraction Failures

Warning

Error: Failed to extract information
Possible Causes:
- PDF does not contain nutritional information.
- Document is completely illegible.
- Network connection interrupted.
- AI service temporarily unavailable.

Solutions:
1. Verify the PDF contains food label data.
2. Try the alternative AI model.
3. Check your internet connection.
4. Ensure PDF is not password-protected or corrupted.

6.1.3 Incomplete or Empty Results

Information

This is normal when:
- Document only contains allergen information (nutritional values will be empty).
- Document only contains nutritional data (fewer allergens detected).
- Specific nutrients are not listed on the label.
- Text is too blurry or small to read accurately.

6.1.4 Low Confidence Results

Tip

1. Re-scan at higher quality (300+ DPI).
2. Try the alternative AI model.
3. Verify results manually.
4. Use better lighting if photographing a physical label.
5. Crop the PDF to include only the nutrition label area.

6.2 Browser Compatibility

Browser	Supported	Notes
Chrome/Edge	✓	Recommended, full features
Firefox	✓	Recommended, full features
Safari	✓	Use file picker for upload
Mobile Chrome	✓	Upload from camera or files
Mobile Safari	✦	Limited: use file picker only

Warning

Safari users: Use the "Click to browse" option instead of drag & drop.

7 Frequently Asked Questions

7.1 General Questions

Q: Is NutriScanner free to use?

A: The application itself may be free to access. AI processing can require API credits. Check your deployment details.

Q: Do I need to create an account?

A: Not necessarily; this depends on your deployment and access settings.

Q: Is my data private and secure?

A: PDFs are processed per deployment policy. Consult your administrator for data-retention and privacy details.

Q: Can I process multiple PDFs at once?

A: Typically one PDF at a time; batch support depends on implementation.

7.2 Technical Questions

Q: What's the difference between the AI backends?

A: Some backends prioritize accuracy and robustness on low-quality inputs; others prioritize speed and cost-effectiveness. Choose according to your needs.

Q: Why does extraction take different amounts of time?

A: Processing time depends on document complexity, size, quality, and the selected AI backend.

Q: Can I upload images instead of PDFs?

A: Convert images to PDF first (many phones can save or print to PDF). The app accepts PDF files.

7.3 Results and Accuracy

Q: How accurate is NutriScanner?

A: Depends on input quality:

- High confidence: typically >95% accuracy.
- Medium confidence: 85–95% accuracy.
- Low confidence: 70–85% accuracy — verify manually.

Q: Can I export or save the results?

A: Many deployments support copy, screenshot, or export. Check the application for CSV/Excel export options.

8 Supported Languages

8.1 Hungarian Support

Allergens:

- Glutén
- Tojás
- Rák
- Hal
- Földimogyoró
- Szója
- Tej
- Diófélék
- Zeller
- Mustár

Nutrition Terms:

- Energia
- Zsír
- Szénhidrát
- Cukor
- Fehérje
- Nátrium
- Tápértékadatok
- Átlagos tápérték

8.2 English Support

Common Terms:

- Nutrition Facts
- Ingredients
- Contains
- May contain traces
- Per 100g
- Per serving

Nutritional Values:

- Energy
- Total Fat
- Carbohydrates
- Of which sugars
- Protein
- Salt / Sodium

8.3 Bilingual Documents

- Automatically detects both languages when present.
- Extracts data from either language section.
- Prioritizes clearer or more complete information.

9 Appendix

9.1 EU-Regulated Allergens Reference

Allergen	Includes
Gluten	Wheat, barley, rye, oats, spelt, kamut
Egg	Egg white, egg yolk, egg protein, albumin
Crustaceans	Shrimp, crab, lobster, prawns, crayfish
Fish	All fish species, fish gelatin
Peanut	Peanuts, peanut butter, peanut oil
Soy	Soybeans, soy protein, soy lecithin, tofu
Milk	Milk, cheese, butter, cream, whey, lactose
Tree Nuts	Almonds, walnuts, cashews, pistachios, hazelnuts, pecans
Celery	Celery stalks, celery root, celeriac, celery seeds
Mustard	Mustard seeds, mustard powder, mustard sauce

9.2 Nutritional Units Reference

kJ Kilojoules - metric energy unit.

kcal Kilocalories (Calories) - common energy unit.

g Grams - mass measurement.

mg Milligrams - 1/1000 of a gram.

per 100g Values for 100 grams of product.

per 100ml Values for 100 milliliters of product.

per serving Values for one typical serving size.

Information

NutriScanner preserves original units exactly as they appear in the document. No conversions are performed automatically.

9.3 Quick Reference Card

Action	Instructions
Upload PDF	Drag & drop or click to browse (max 5MB)
Select Model	Choose the desired AI backend
Extract Data	Click "Extract Information" button
View Progress	Watch progress bar and status messages
Check Results	Review allergens, nutrition, confidence
Scan Another	Click "Scan Another" button
Switch Models	Select different model before uploading
Get Help	Click "Documentation" or contact support

9.4 Glossary

AI Model Artificial Intelligence system that reads and extracts data.

Base64 Encoding method used sometimes for file uploads.

Confidence Level Assessment of extraction reliability (High/Medium/Low).

OCR Optical Character Recognition - reading text from images.

Streaming Real-time data processing with live progress updates.

Vision AI AI systems that interpret images and documents.

10 Contact & Support

10.1 Getting Help

If you encounter issues or need assistance:

1. Check this manual for troubleshooting steps.
2. Review the FAQ section for common questions.
3. Visit the documentation page within the application.
4. Contact your administrator for technical support.

10.2 Reporting Issues

When reporting a problem, please include:

- What you were trying to do.
- What happened instead.
- Error messages (if any).
- Browser and operating system.
- Document type (digital PDF or scan).
- AI model selected.

10.3 Feature Requests

We're constantly improving NutriScanner! If you have ideas:

- Submit feedback through the application.
- Describe the feature and how it would help you.
- Explain your use case.

10.4 Additional Resources

- **GitHub Repository:** `https://github.com/purelyricky/nutriscan`
- **Developer Documentation:** Available at the application's /docs.
- **Source Code:** Open source and available for review.

Thank you for using NutriScanner!

Making nutritional information accessible through AI

Version 1.0 — October 26, 2025