

Package: soccerrsr (via r-universe)

July 20, 2026

Type Package

Title SOCcer Occupation and Industry Classification

Version 0.1.1

Description Brings SOCcer and CLIPS from online only to your favorite working environment - R. Thanks to the power of Rust and ONNX.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

Roxygen list(markdown = TRUE)

Config/rextendr/version 0.4.2

Config/rextendr/use_symbols FALSE

SystemRequirements Cargo (Rust's package manager), rustc >= 1.65.0, xz

Depends R (>= 4.2)

Config/pak/sysreqs xz-utils libclang-dev

Repository <https://danielruss.r-universe.dev>

Date/Publication 2026-07-16 21:03:39 UTC

RemoteUrl <https://github.com/danielruss/soccerrsr>

RemoteRef main

RemoteSha d597a8e86a766a07b38ddcf1a0938bc3b47a0914

Contents

clips	2
embed	2
embed_job	3
soccer_net	3
SOCcerNET	4

Index	6
--------------	----------

clips	<i>Run clips</i>
-------	------------------

Description

Run clips

Usage

```
clips(df, n, block_size)
```

embed	<i>Embed job descriptions or product/service descriptions</i>
-------	---

Description

Generates dense vector embeddings using the GIST-small-Embedding model. Automatically detects the type of input based on column names.

Usage

```
embed_job(text1, text2 = NULL)

embed_jobs(df)
```

Arguments

- | | |
|-------|---|
| text1 | A single character string to embed (job title or product/service) |
| text2 | An optional single character string for additional context (job task) |
| df | A data frame containing either: <ul style="list-style-type: none">• products_services — for CLIPS data• JobTitle — for SOCcerNET data• JobTask optional - only if JobTitle is provided |

Details

NA values in JobTitle, products_services, or JobTask will be replaced with a single space character and a warning will be issued indicating which rows are affected.

Value

A numeric vector of length 384 (embed_job) or a numeric matrix of shape n x 384 (embed_jobs) where each row is an embedding.

See Also[SOCcerNET](#)**Examples**

```
## Not run:
# Single embedding
embed_job("Doctor", "Diagnose patients")
embed_job("Doctor")

# Batch embedding - SOCcerNET
df <- data.frame(JobTitle = c("Doctor", "Lawyer"), JobTask = c("Diagnose patients", "Draft contracts"))
embed_jobs(df)

# Batch embedding - CLIPS
df <- data.frame(products_services = c("Software development", "Cloud hosting"))
embed_jobs(df)

## End(Not run)
```

`embed_job`*Embed a job*

Description

Embed a job

Usage

```
embed_job(text1, text2 = NULL)
```

`soccer_net`*Run SOCcerNET from R*

Description

Run SOCcerNET from R

Usage

```
soccer_net(df, n, block_size)
```

SOCcerNET

Occupation or Industry coding

Description

Classifies job descriptions or product/service descriptions into occupational or industry codes. `run_soccernet()` codes jobs into US SOC 2010 codes. `run_clips()` codes products and services into industry classifications. SOCcerNET can use optional `soc1980`, `noc2011`, or `isco1988` columns to assist in the `soc2010` coding. CLIPS can use a `sic1987` columns.

Usage

```
run_soccernet(df, n = 10, block_size = 8)
```

```
run_clips(df, n = 10, block_size = 20)
```

Arguments

<code>df</code>	A data frame containing the text to classify. See Details for required columns.
<code>n</code>	Number of candidate codes to return per row (default: 10)
<code>block_size</code>	Number of rows to process per batch, tune for performance (default: 8 for SOCcerNET, 20 for CLIPS)

Details

Required columns:

`run_soccernet()` requires:

- `JobTitle` — job title text (*required*)
- `JobTask` — job task description (*optional, created as empty string if missing*)
- `Id` — unique row identifier (*optional, auto-generated as row-001, row-002, etc. if missing*)

Optionally, crosswalk columns `soc1980`, `noc2011`, and `isco1988` can be provided as lists or character vectors.

`run_clips()` requires:

- `products_services` — product or service description (*required*)
- `Id` — unique row identifier (*optional, auto-generated if missing*)

Performance:

`block_size` controls how many rows are processed at once. Larger values may improve throughput but use more memory. Tune based on your hardware.

Value

The input data frame with additional columns containing the top `n` classification codes and their scores, joined by `Id`.

See Also[embed_jobs\(\)](#)**Examples**

```
## Not run:
# SOCcerNET
df <- data.frame(
  JobTitle = c("Staff Scientist", "Data Engineer"),
  JobTask  = c("Develop chemical assays", "Build data pipelines")
)
run_soccernet(df)
run_soccernet(df, n = 5, block_size = 16)

# CLIPS
df <- data.frame(products_services = c("Software development", "Cloud hosting"))
run_clips(df)

## End(Not run)
```

Index

CLIPS (SOCcerNET), [4](#)

clips, [2](#)

clips (SOCcerNET), [4](#)

embed, [2](#)

embed_job, [3](#)

embed_job (embed), [2](#)

embed_jobs (embed), [2](#)

embed_jobs(), [5](#)

run_clips (SOCcerNET), [4](#)

run_soccernet (SOCcerNET), [4](#)

soccer_net, [3](#)

SOCcerNET, [3](#), [4](#)

soccernet (SOCcerNET), [4](#)