



1099-R Labeling Instructions

Evaluation Set Requirements



In order to do fair and thorough evaluation of the model, the evaluation set should take the following dimensions into consideration

1. **Revisions:** Documents should be in English for years 2019 & 2020.
2. **Size:** It is recommended to have 500+ documents for each evaluation set.
3. **Template Diversity:** The documents should be diverse, e.g. different templates, issue year, different sources, states (if applicable) ... etc.
4. **Value Diversity:** Documents should capture all possible values.
5. **Handwritten:** It is preferred to separate handwritten documents from typed ones (i.e. have 2 separate evaluation sets) and report metrics for each.

Processor Version



These instructions are specific for 1099-R Processor version:

“Google Release Candidate (Default) pretrained-1099r-v2.0-2022-07-25”

PROCESSOR DETAILS TRAIN **EVALUATE & TEST** MANAGE VERSIONS

Version

Filter | type to filter

- Google Release Candidate (Default)**
pretrained-1099r-v2.0-2022-07-25 | Not evaluated
- Google Stable
pretrained-1099r-v1.0-2021-08-11 | Not evaluated

1 No test dataset exists for this processor yet. You need to annotate a test dataset in order to evaluate this version.

All labels

| | |
|---------------------|---|
| F1 score ? | - |
| Precision ? | - |
| Recall ? | - |
| Last evaluated | - |
| Test documents | - |
| Evaluated documents | - |
| Invalid documents ? | - |
| Failed documents ? | - |

Entity List

| Field | Description | Base Type |
|--------------------------|---|-----------|
| FormYear | 1099-R Form Year | string |
| PayerFirstName | First Name of the Payer | string |
| PayerLastName | Last Name of the Payer | string |
| PayerMiddleInitial | Middle Initial of the Payer | string |
| PayerOrganizationName | Organization Name of the Payer | string |
| PayerStreetAddress_Line1 | Enter Street Address line 1 of the Payer | string |
| PayerStreetAddress_Line2 | Enter Street Address line 2 of the Payer | string |
| PayerCity | Enter City of the Payer (City or Town) | string |
| PayerState | Enter State of the Payer(State or Province) | string |
| PayerZipcode | Enter the Zipcode of the Payer | string |
| RecipientFirstName | First Name of the Recipient | string |
| RecipientLastName | Last Name of the Recipient | string |

Entity List

| Field | Description | Base Type |
|------------------------------|---|-----------|
| RecipientMiddleInitial | Middle Initial of the Recipient | string |
| RecipientOrganizationName | Organization Name of the Recipient | string |
| RecipientStreetAddress_Line1 | Enter Street Address line 1 of the Recipient (Including Apt No.) | string |
| RecipientStreetAddress_Line2 | Enter Street Address line 2 of the Recipient | string |
| RecipientCity | Enter City of the Recipient (City or Town) | string |
| RecipientState | Enter State of the Recipient (State or Province) | string |
| RecipientZipcode | Enter the Zipcode of the Recipient | string |
| RecipientTIN | Enter Recipient Tax Identification Number | string |
| FederalIncomeTaxWithheld | Total amount of federal tax withheld during that year | money |
| GrossDistribution | Total amount paid to retiree in that year before taxes or other deductions – does not include health insurance reimbursements | money |
| TaxableAmount | Enter the Taxable Amount | money |
| DistributionCode | DistributionCode Identifies the type of distribution received | string |

Entity List

| Field | Description | Base Type |
|----------------------------------|--|-----------|
| Other | other | string |
| TotalEmployeeContributions | Enter Total Employee Contributions | money |
| AmountAllocableToIRRWithin5Years | Enter Amount Allocable To IRR Within 5 Years | money |
| StateTaxWithheld_Line1 | Amounts deducted from disbursements for state income tax purposes. | money |
| StateTaxWithheld_Line2 | Amounts deducted from disbursements for state income tax purposes. | money |
| LocalTaxWithheld_Line1 | Enter the amount of Local tax withheld from the distribution. | money |

Note: There will be zero or one instance of each entity.

Entity Annotation .

PayerOrganizationName

PayerMiddleInitial

PayerLastName

PayerZipcode

RecipientTIN

GrossDistribution

PayerFirstName

PayerStreetAddress_Line1

PayerStreetAddress_Line2

PayerCity

PayerState

RecipientOrganizationName

RecipientFirstName

RecipientMiddleInitial

RecipientLastName

RecipientStreetAddress_Line1

RecipientStreetAddress_Line2

RecipientCity

RecipientState

RecipientZipcode

| | | | | | | |
|---|--|--|--|--|---|--|
| <div><input type="checkbox"/> CORRECTED (if checked)</div> <div>PAYER'S name, street address, city or town, state or province, country, ZIP or foreign postal code, and phone no.</div> <div>Huber LLC</div> <div>Diaz P Isle</div> <div>121 Worchester Rd</div> <div>Apt 101</div> <div>Framingham MA 17014</div> <div>PAYER'S TIN</div> <div>RECIPIENT'S TIN</div> <div>138-05-2338</div> | | | <div>1 Gross distribution</div> <div>OMB No. 1545-0119</div> <div>2020</div> <div>2a Taxable amount</div> <div>Form 1099-R</div> | | <div>Distributions From Pensions, Annuities, Retirement or Profit-Sharing Plans, IRAs, Insurance Contracts, etc.</div> <div>FormYear</div> <div>TaxableAmount</div> | |
| <div>not determined <input type="checkbox"/></div> <div>Total distribution <input type="checkbox"/></div> | | | <div>3 Capital gain (included in box 2a)</div> <div>4 Federal income tax withheld</div> <div>11028.04</div> | | <div>Copy B</div> <div>Report this income on your federal tax return. If this form shows federal income tax withheld in box 4, attach this copy to your return.</div> <div>FederalIncomeTaxWithheld</div> <div>DistributionCode</div> | |
| <div>RECIPIENT'S name</div> <div>Perez-Group</div> <div>Moran M Plaza</div> <div>Street address (including apt. no.)</div> <div>504 W. Center St</div> <div>Apt 104</div> <div>City or town, state or province, country, and ZIP or foreign postal code</div> <div>Aliquippa PA 15001</div> | | | <div>5 Employee contributions/ Designated Roth contributions or insurance premiums</div> <div>6 Net unrealized appreciation in employer's securities</div> | | <div>Other</div> <div>TotalEmployeeContributions</div> | |
| <div>10 Amount allocable to IRR within 5 years</div> <div>11 1st year of desig. Roth contrib.</div> <div>12 FATCA filing requirement</div> <div>14 State tax withheld</div> <div>15 State/Payer's state no.</div> <div>16 State distribution</div> | | | <div>7 Distribution code(s)</div> <div>IRA/ SEP/ SIMPLE</div> <div>8 Other</div> <div>9a Your percentage of total distribution</div> <div>9b Total employee contributions</div> <div>17 Local tax withheld</div> <div>18 Name of locality</div> <div>19 Local distribution</div> | | <div>This information is being furnished to the IRS.</div> <div>StateTaxWithheld_Line1</div> <div>StateTaxWithheld_Line2</div> <div>LocalTaxWithheld_Line1</div> | |
| <div>Account number (see instructions)</div> <div>13 Date of payment</div> <div>17 Local tax withheld</div> <div>18 Name of locality</div> <div>19 Local distribution</div> | | | <div>233.50</div> <div>677.65</div> <div>8</div> <div>234.56</div> <div>12273.89</div> <div>9158.27</div> <div>5593.58</div> <div>9291.63</div> <div>2963.45</div> | | | |

Form 1099-R

www.irs.gov/Form1099R

Department of the Treasury - Internal Revenue Service



General Labeling Instructions

Occurrence Type



There are 4 occurrence types which restricts the number of annotations for each entity:

- **OPTIONAL_ONCE**: Entity can be present zero or multiple times but should have one unique value.
- **OPTIONAL_MULTIPLE**: Entity will appear zero or multiple times and with different values.
- **REQUIRED_ONCE**: Entity can be present once or multiple times but should have one unique value.
- **REQUIRED_MULTIPLE**: Entity will appear once or multiple times and with different values.

All 1099-R form entities belongs to any of the occurrence type mentioned above .

Optical Character Recognition (OCR) Errors



Drawing bounding boxes around entities will automatically populate a text captured through OCR (Optical Character Recognition). In some cases, the captured text is not accurate and doesn't reflect what is on the page, especially for documents that are blurry or of low resolution. These values should be corrected manually through "Value" input text.

Bounding Box Annotation

Bounding boxes should fit the correct text as much as possible:

- Avoid overlapping bounding boxes. This can be tricky when values are close to each other.
- Avoid oversized boxes which incorporate background artifacts such as table borders.

Example 1

A screenshot of a Form 1099-NEC with two red bounding boxes. One box is around the value '\$ 141.67' and the other is around '\$ 762.58'. The two boxes overlap, which is incorrect.

| | | |
|-------|--------------------------|---------------|
| | <input type="checkbox"/> | |
| | 5 State tax withheld | 6 State/Payer |
| | \$ 141.67 | |
| | \$ 762.58 | |
| ords) | www.irs.gov/Form1099NEC | Departmer |

Incorrect Annotation:
Bounding boxes are overlapping

Example 2

A screenshot of a Form 1099-NEC with two red bounding boxes. One box is around the value '\$ 141.67' and the other is around '\$ 762.58'. The two boxes are separate and do not overlap, which is correct.

| | | |
|-------|--------------------------|---------------|
| | <input type="checkbox"/> | |
| | 5 State tax withheld | 6 State/Payer |
| | \$ 141.67 | |
| | \$ 762.58 | |
| ords) | www.irs.gov/Form1099NEC | Departmer |

Correct Annotation

Example 3

A screenshot of a Form 1099-NEC with two red bounding boxes. One box is around the value '\$ 141.67' and the other is around '\$ 762.58'. The boxes are oversized and include background artifacts like table borders, which is incorrect.

| | | |
|-------|--------------------------|---------------|
| | <input type="checkbox"/> | |
| | 5 State tax withheld | 6 State/Payer |
| | \$ 141.67 | |
| | \$ 762.58 | |
| ords) | www.irs.gov/Form1099NEC | Departmer |

Incorrect Annotation:
Bounding boxes oversized

Case Sensitivity



DocAI evaluation is case sensitive and can affect the final evaluation. While annotation please be aware to fix any casing issues generated by OCR.

Please note that Entity values are also case sensitive and the values should be fixed

Examples:

- "Huber llc" != "Huber LLC"
- "Diaz p" != "Diaz P"

Money

For money entries the annotation bounding box and value text should include the currency if present on the document.

If the document contains currency with no values, then annotations could be omitted, as shown in the example on the bottom right.

Contract Summary

| | This Quarter | Year To Date | Since Issue Date ¹ |
|-------------------------------|---------------|---------------|-------------------------------|
| Beginning Value on 09/28/2018 | \$201,931.64 | | |
| Total Premium | \$0.00 | \$4,000.00 | \$146,388.89 |
| Total Withdrawals | \$0.00 | \$0.00 | \$0.00 |
| Total Tax Withheld | \$0.00 | \$0.00 | |
| Net Change | (\$23,104.00) | (\$15,683.48) | |
| Ending Value on 12/31/2018 | \$178,827.64 | | |

| | | | |
|--|-----------------------------|----------------------------------|-----------------------------|
| FATCA filing requirement <input type="checkbox"/> | 12 State tax withheld \$ | 13 State/Payer's state no. \$ | 14 State distribution \$ |
| | | | |
| Date of Payment | 15 Local tax withheld \$ | 16 Name of locality \$ | 17 Local distribution \$ |
| | | | |

www.irs.gov/form 1099r Department of the Treasury - Internal Revenue Service

Empty Values

In general, if the value does not exist in the form then there is no need to label it. In the example below `FederalIncomeTaxWithheld` should not be annotated since there is no value present in the designated field.

It is also possible to annotate empty entities (with empty values) but that will not affect the final metrics.

| | | | | | | |
|--|--|-------------------------------------|---|-----------------------|--|----------------------------|
| 42 Fairhaven Commons Way, Fairhaven MA 2719 | | | \$ | | Form 1099-R | |
| PAYER'S TIN | | RECIPIENT'S TIN | 2b Taxable amount not determined <input type="checkbox"/> | | Total distribution <input type="checkbox"/> | |
| 060151575 | | 123-45-4351 | 3 Capital gain (included in box 2a) | | 4 Federal income tax withheld | |
| RECIPIENT'S name | | | 5 Employee contributions/ Designated Roth contributions or insurance premiums | | 6 Net unrealized appreciation in employer's securities | |
| Forks player | | | \$ | | \$ | |
| Street address (including apt. no.) | | | 7 Distribution code(s) | | 8 Other | |
| 236 Apt. Common street, Fairhaven MA 2719 | | | 7 | | % | |
| City or town, state or province, country, and ZIP or foreign postal code | | | 9a Your percentage of total distribution % | | 9b Total employee contributions | |
| 10 Amount allocable to IRR within 5 years | | 11 1st year of desig. Roth contrib. | 12 FATCA filing requirement <input type="checkbox"/> | 14 State tax withheld | | 15 State/Payer's state no. |
| \$ | | | | 0.00 | | |
| Account number (see instructions) | | 13 Date of payment | 17 Local tax withheld | | 18 Name of locality | |
| | | | 0.00 | | | |
| | | | 0.00 | | | |

FederalIncomeTaxWithheld