

RandomWalk

Problem Statement:

Machine learning
system to deduplicate
people data



The solution

A web application usable by anyone which generates a new file with clustered duplicate data.

Almabase De-Duplicator

Upload Data

Name :

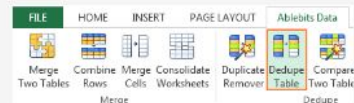
Description :

Upload File

Choose File No file chosen

Submit

Upload Format



FILE	HOME	INSERT	PAGE LAYOUT	Ablebits Data
Merge Two Tables	Combine Rows	Merge Cells	Consolidate Worksheets	Duplicate Remover
				Deduplicate Table
				Compare Two Tables
				Deduplicate

A2 X ✓ fx 100023

1	Order no.	Order date	Item	Tax	Delivery
2	100023		42373 Apples	19%	Delivered
3	100028		42373 Apples		In transit

The file must contain one row for every record, with first row indicating the name of each column.

Almbase De-Duplicator

Dhruv

127.0.0.1:8000/page1/7/

AppsCSMLGWSPOGAIKlinYouTubeFacebookTwitterDiscordMailGoogleDriveTrelloZP

Other Bookmarks

Almbase De-Duplicator

1. Denne Fields

Select columns which you think tell most information about each row in your spreadsheet and how would you like to compare them.

Field in Project

Address

Field in Project

Name

Field in Project

Phone

+ Add more

Almbase De-Duplicator

127.0.0.1:8000/page2/7/

AppsCSMLGWSPOGAIKlinYTfbTWZPTZP

Dhruv

Almbase De-Duplicator

2. Train Model

Below are 2 records, **Record A** and **Record B**.

Field	Record A	Record B
Address	5423 w diversey	600 n. leavitt
Name	onward neighborhood house - belmont-cragin	onward neighborhood house
Phone	None	6666726

Do these records represnt the same thing?

Yes

No

Finish

Cluster ID	confidence_score	Id	Name	Ad
0	0.999847263098	0	Salvation Army - Temple / Salvation Army	1 f
0	0.999847263098	1	Salvation Army - Temple / Salvation Army	1 f
1	0.999448135495	2	National Louis University - Dr. Effie O. Ellis Training Center	10
1	0.999448135495	3	National Louis University - Dr. Effie O. Ellis Training Center	10
3	0.999980866909	4	Board Trustees-City Colleges of Chicago - Olive-Harvey College	10
3	0.999980866909	5	Board Trustees-City Colleges of Chicago - Olive-Harvey College	10
4	0.999980866909	6	Easter Seals Society of Metropolitan Chicago - Gilchrist-Marchman	10
4	0.999980866909	7	Easter Seals Society of Metropolitan Chicago - Gilchrist-Marchman	10
5	0.999981	8	Hull House Association - Uptown Head Start / Hull House	10
6	0.999980866909	9	Hull House Association - Child Dev. Central Office- Hull House CDS-Homes	10
6	0.999980866909	10	Hull House Association - Child Dev. Central Office- Hull House CDS-Homes	10
7	0.999980866909	11	Catholic Charities of the Archdiocese of Chicago - St. Kevin	10
7	0.999980866909	12	Catholic Charities of the Archdiocese of Chicago - St. Kevin	10
94		13	Chicago State University - Community Learning Center	10
8	0.999981	14	Catholic Charities of the Archdiocese of Chicago - Cordia Marian	11
8	0.999981	15	Catholic Charities of the Archdiocese of Chicago - Cordia Marian	11
95		16	Thresholds - Thresholds Mothers' Project	11