

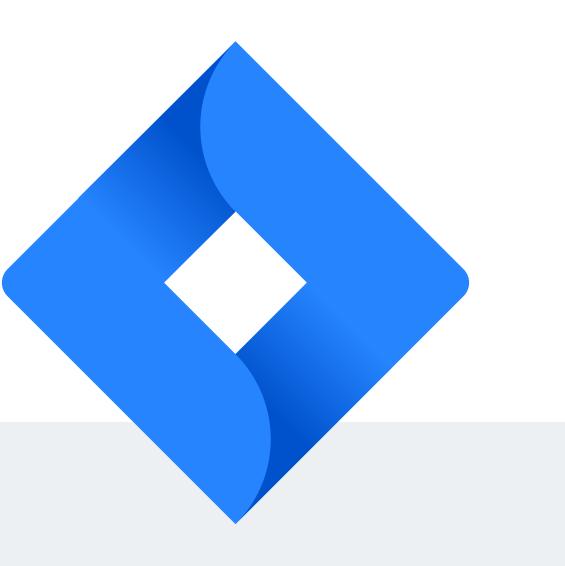
Salary Prediction



Table of Content

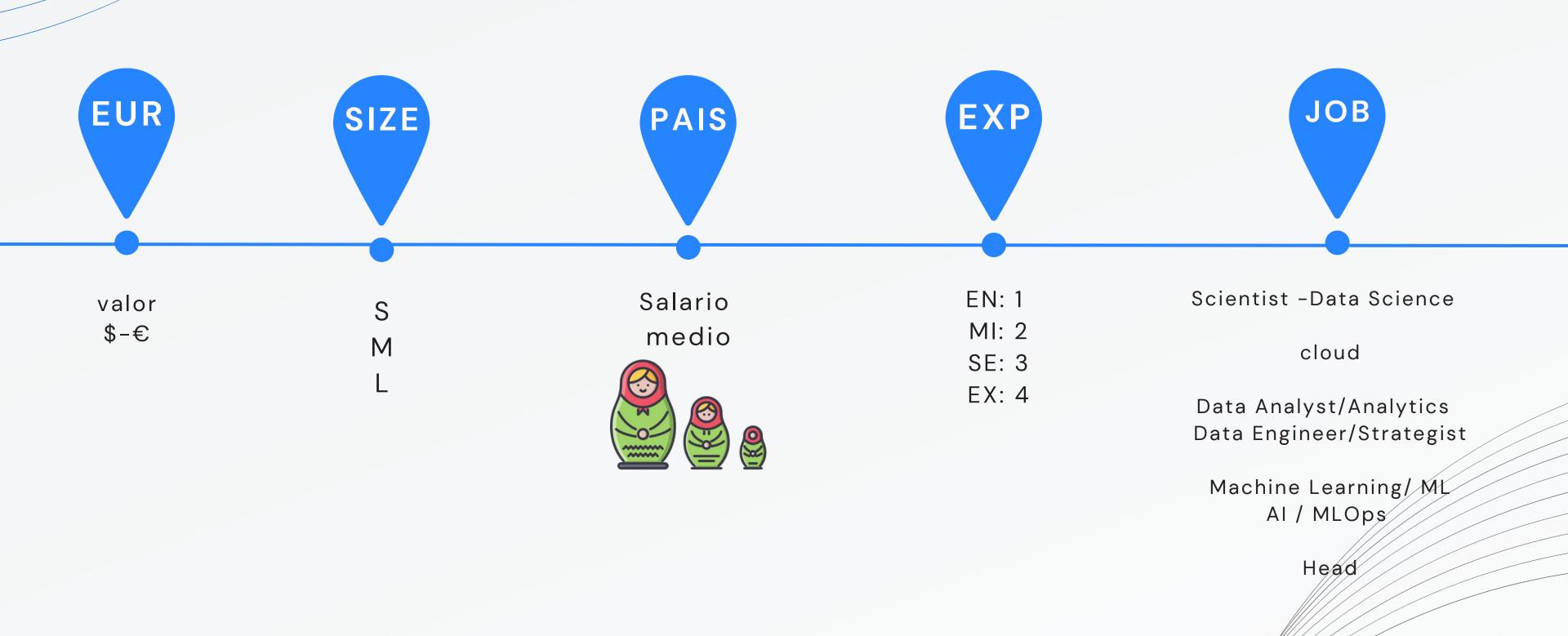
- Datos originales
- Data engineering
- Modelos propuestos
- Modelo final
- Streamlit

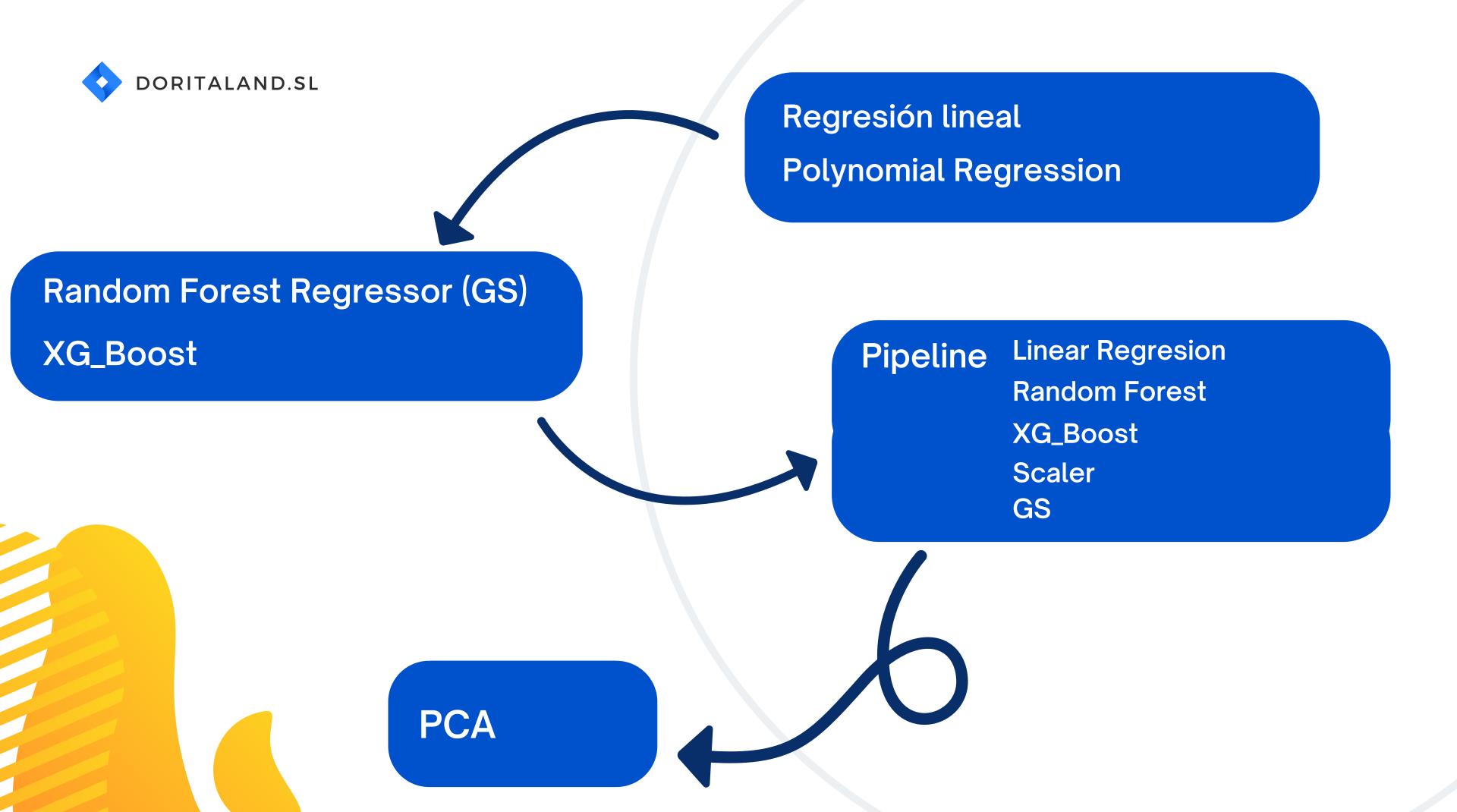
	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
0	2023	SE	FT	Principal Data Scientist	80000	EUR	85847	ES	100	ES	L
1	2023	MI	CT	ML Engineer	30000	USD	30000	US	100	US	S
2	2023	MI	CT	ML Engineer	25500	USD	25500	US	100	US	S
3	2023	SE	FT	Data Scientist	175000	USD	175000	CA	100	CA	М
4	2023	SE	FT	Data Scientist	120000	USD	120000	CA	100	CA	М



<class 'pandas.core.frame.DataFrame'> RangeIndex: 3755 entries, 0 to 3754 Data columns (total 11 columns): Non-Null Count Dtype Column work_year 3755 non-null int64 experience_level 3755 non-null object employment_type object 3755 non-null job_title object 3755 non-null 3755 non-null salary int64 salary_currency object 3755 non-null salary_in_usd 3755 non-null int64 employee_residence 3755 non-null object 8 remote_ratio 3755 non-null int64 company_location 3755 non-null object 10 company_size object 3755 non-null dtypes: int64(4), object(7) memory usage: 322.8+ KB







	Linear	Poly	RForest	GXBoost	Pipe	PCA
Score	0.33	10696613916437	0.35	0.32	0.37	0.35
MAE	35590	15558563751	34166	35473	34132	34547







NEXT STEPS





Streamlit



Nuevos data engineering



Nueva recolección de data

GRACIAS POR SU ATENCION