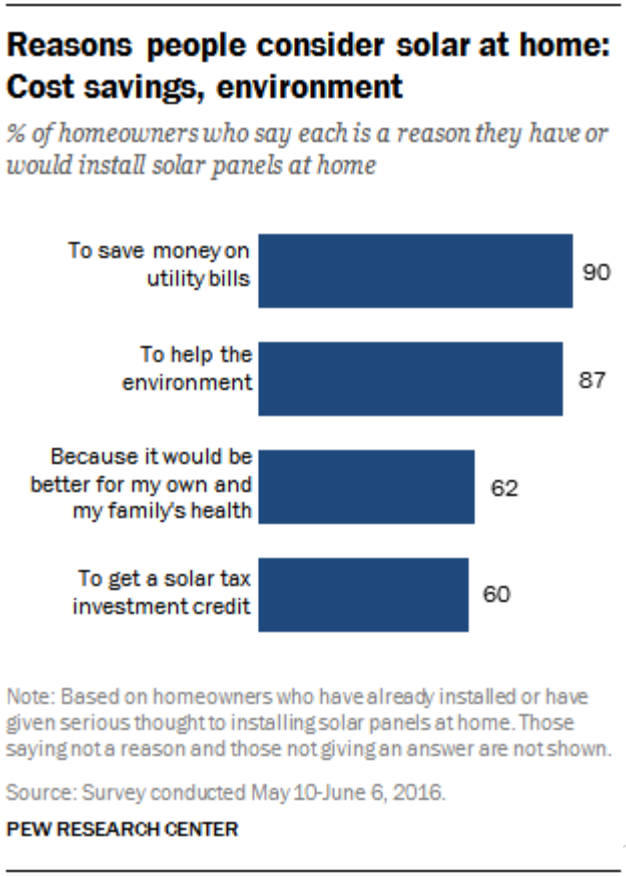
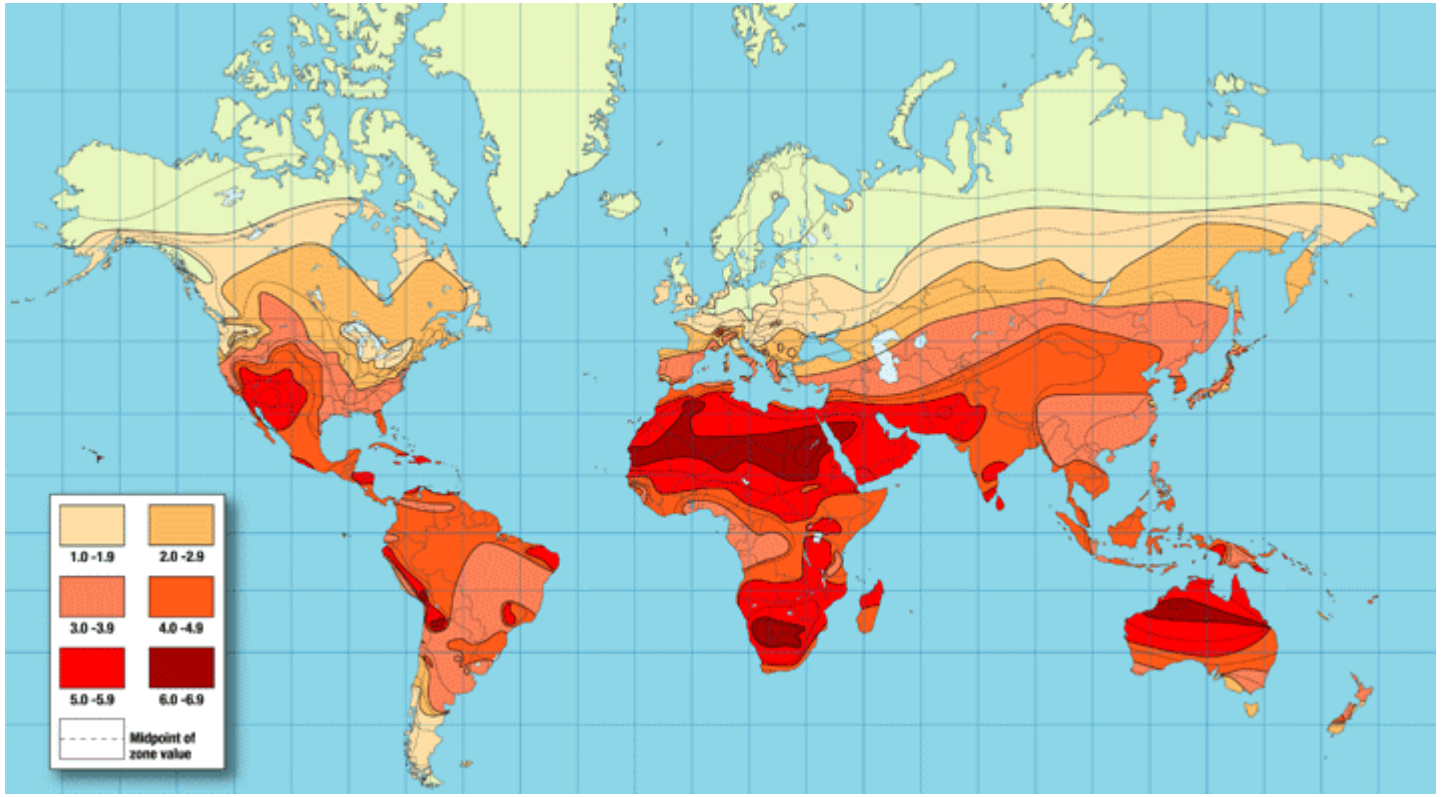
Predicting solar installation rates in different locations using demographic data

Solar installations are integral in our transition to renewable energy sources

- As the effects of global climate change are becoming better understood, the move towards renewable energy sources has become an important focus
- Solar energy is emerging as one of the most popular forms of renewable energy for reasons such as decreasing costs, environmental ethics, health, government incentives and accessibility



Australia has some of the world's best solar exposure



Prediction has value to policy makers, investors and researchers

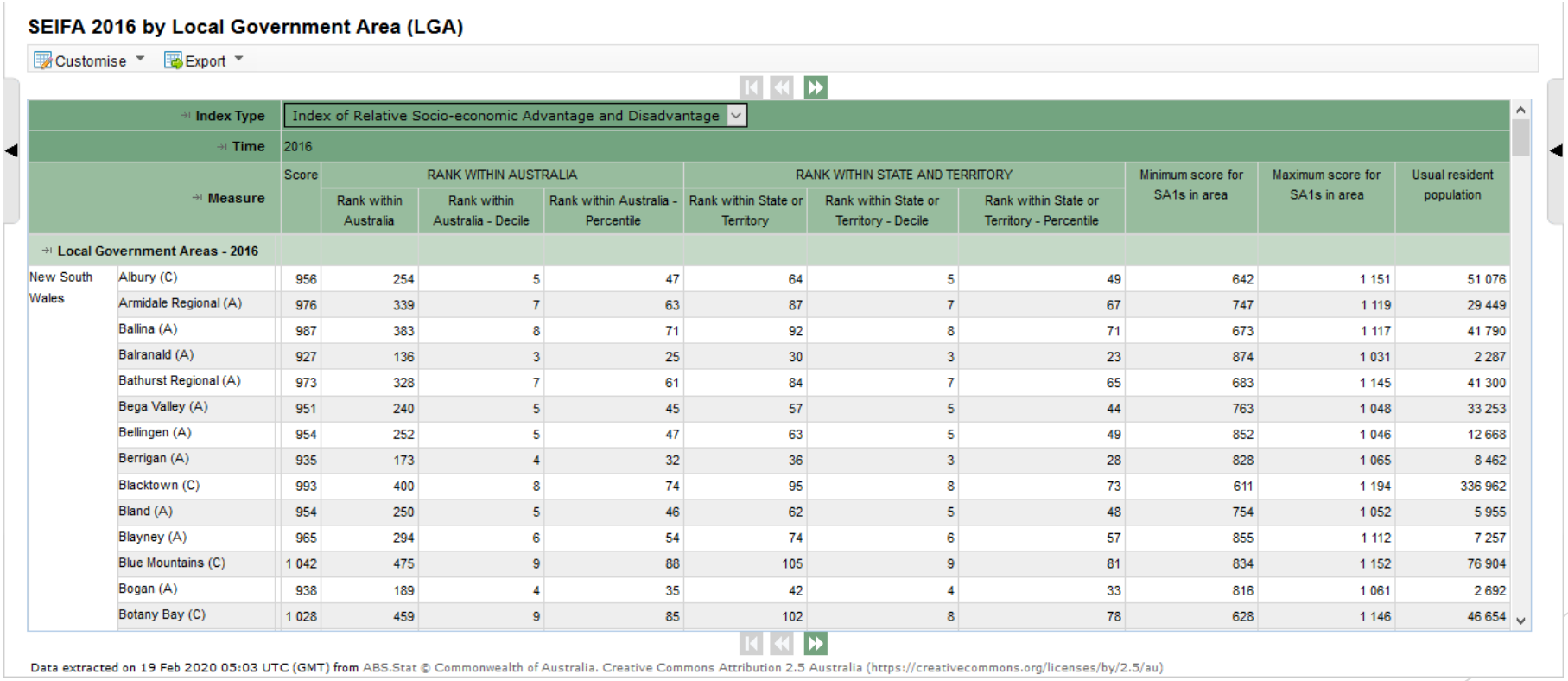
- Reliable predictions on installation rates may guide projections and highlight both the level and content of policy intervention justified in affecting solar uptake
- Investors in the solar industry hold an advantage if they have reliable guidance on which local government areas in the future (accounting for demographic shifts) are likely to install solar
- Results from attempting to answer this question may lead to new lines of inquiry among researchers (e.g. if we can indeed predict installation rates then that provides justification to then do further research on which exact demographic factors have an effect)

Scope of analysis

- To answer: “Can we predict solar installation rates in the different local government areas of Australia using demographic data?”
- To build a machine learning algorithm to make predictions on solar installation rates in different local government areas (LGAs) given certain demographic factors
- (Not included in analysis): which specific demographic factors are related to different solar installation rates

Solar installation rate data from Australian Photovoltaic Institute

Demographic information datasets from Australian Bureau of Statistics



SEIFA 2016 by Local Government Area (LGA)

Customise Export

Index Type		Index of Relative Socio-economic Advantage and Disadvantage									
Time		2016									
Measure		Score	RANK WITHIN AUSTRALIA			RANK WITHIN STATE AND TERRITORY			Minimum score for SA1s in area	Maximum score for SA1s in area	Usual resident population
			Rank within Australia	Rank within Australia - Decile	Rank within Australia - Percentile	Rank within State or Territory	Rank within State or Territory - Decile	Rank within State or Territory - Percentile			
Local Government Areas - 2016											
New South Wales	Albury (C)	956	254	5	47	64	5	49	642	1 151	51 076
	Armidale Regional (A)	976	339	7	63	87	7	67	747	1 119	29 449
	Ballina (A)	987	383	8	71	92	8	71	673	1 117	41 790
	Balranald (A)	927	136	3	25	30	3	23	874	1 031	2 287
	Bathurst Regional (A)	973	328	7	61	84	7	65	683	1 145	41 300
	Bega Valley (A)	951	240	5	45	57	5	44	763	1 048	33 253
	Bellingen (A)	954	252	5	47	63	5	49	852	1 046	12 668
	Berrigan (A)	935	173	4	32	36	3	28	828	1 065	8 462
	Blacktown (C)	993	400	8	74	95	8	73	611	1 194	336 962
	Bland (A)	954	250	5	46	62	5	48	754	1 052	5 955
	Blayney (A)	965	294	6	54	74	6	57	855	1 112	7 257
	Blue Mountains (C)	1 042	475	9	88	105	9	81	834	1 152	76 904
	Bogan (A)	938	189	4	35	42	4	33	816	1 061	2 692
	Botany Bay (C)	1 028	459	9	85	102	8	78	628	1 146	46 654

Data extracted on 19 Feb 2020 05:03 UTC (GMT) from ABS.Stat © Commonwealth of Australia. Creative Commons Attribution 2.5 Australia (https://creativecommons.org/licenses/by/2.5/au)

Data converted to form where demographic predictors are columns

LGA	Density (%)	IEO	IER	IRSAD	IRSD	4 year olds enrolled in preschool or in a preschool program (no.)	5 year olds enrolled in preschool or in a preschool program (no.)	Advanced Diploma/Diploma (%)	Agriculture, Environmental and Related Studies (%)	Architecture and Building (%)	...	Flat or apartment: In a one or two storey block	Flat or apartment: In a three storey block	House or flat attached to a shop, office, etc.	Improved home, tent, sleepers out	Not appliacable	Not stated	Semi-detached, row or terrace house, townhouse etc. with : One storey	Semi-detached, row or terrace house, townhouse etc. with : Two or more storeys	Separate house	Total
Albury (C)	18.8	961.0	960.0	956.0	971.0	580.0	224.0	8.5	2.7	6.3	...	563	112	55	7	91	82	3702	586	18127	23464
Armidale Regional (A)	20.3	1015.0	958.0	976.0	980.0	290.0	70.0	7.2	6.6	4.4	...	647	60	39	13	72	99	1036	189	10483	12738
Ballina (A)	36.4	999.0	998.0	987.0	1003.0	480.0	180.0	9.5	3.2	6.7	...	727	198	38	20	48	144	2532	1847	12760	19154
Balranald (A)	15.8	915.0	969.0	927.0	942.0	18.0	10.0	4.4	9.3	3.3	...	55	0	7	5	20	4	28	3	1044	1221
Bathurst Regional (A)	17.9	978.0	993.0	973.0	986.0	382.0	154.0	8.1	3.0	5.7	...	799	17	31	10	63	171	1495	434	14345	17431

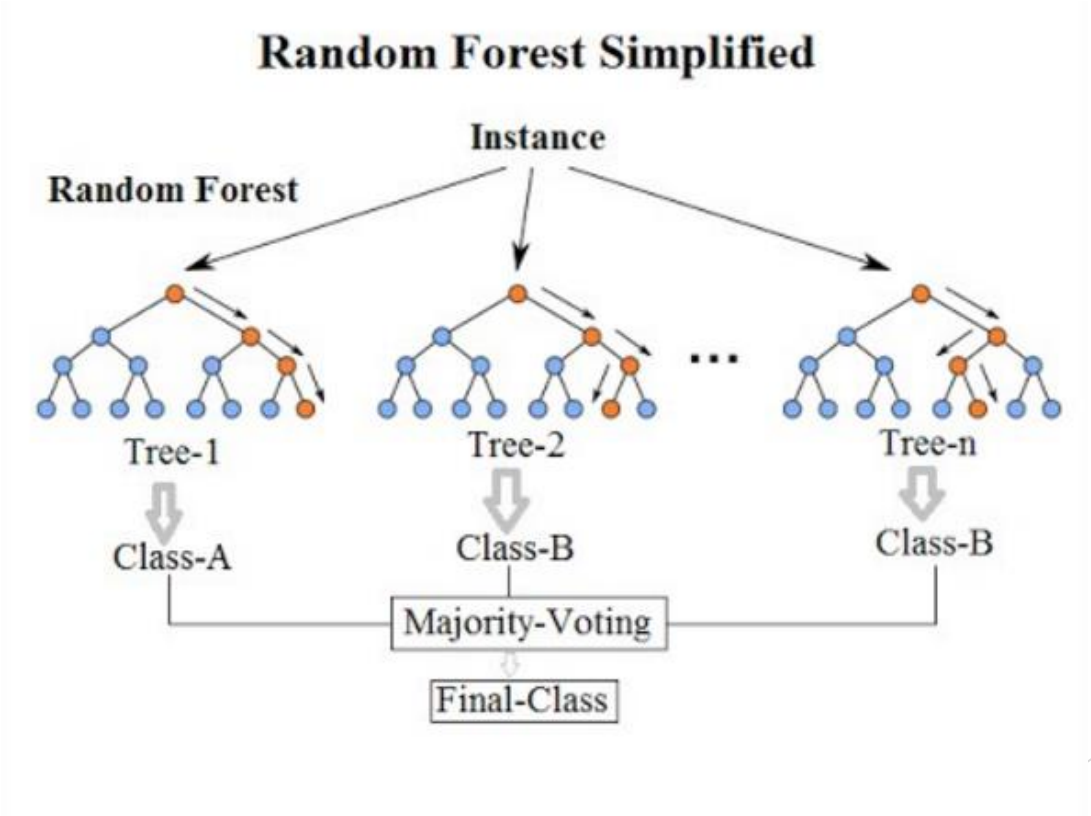
5 rows × 194 columns

Data processing

- Combined dataset size was 481 rows by 194 columns
- Predictors with missing values were imputed or dropped, resulting in dataset size of 481 rows by 125 columns
- Train-test split of 70-30 was performed
- Scaling and principal components transformation of predictors was performed
- The first 20 principal components captured 99.999% of variability in predictors
- These first 20 principal components were used as predictors and the density (solar installation rate) as the response for our machine learning models

Four machine learning models were built

- Multiple Linear Regression Model
- Boosting Regression Model
- Random Forest Model
- Random Forest Model with Hyperparameter Tuning
- Random Forest Model (without tuning) was selected as final model

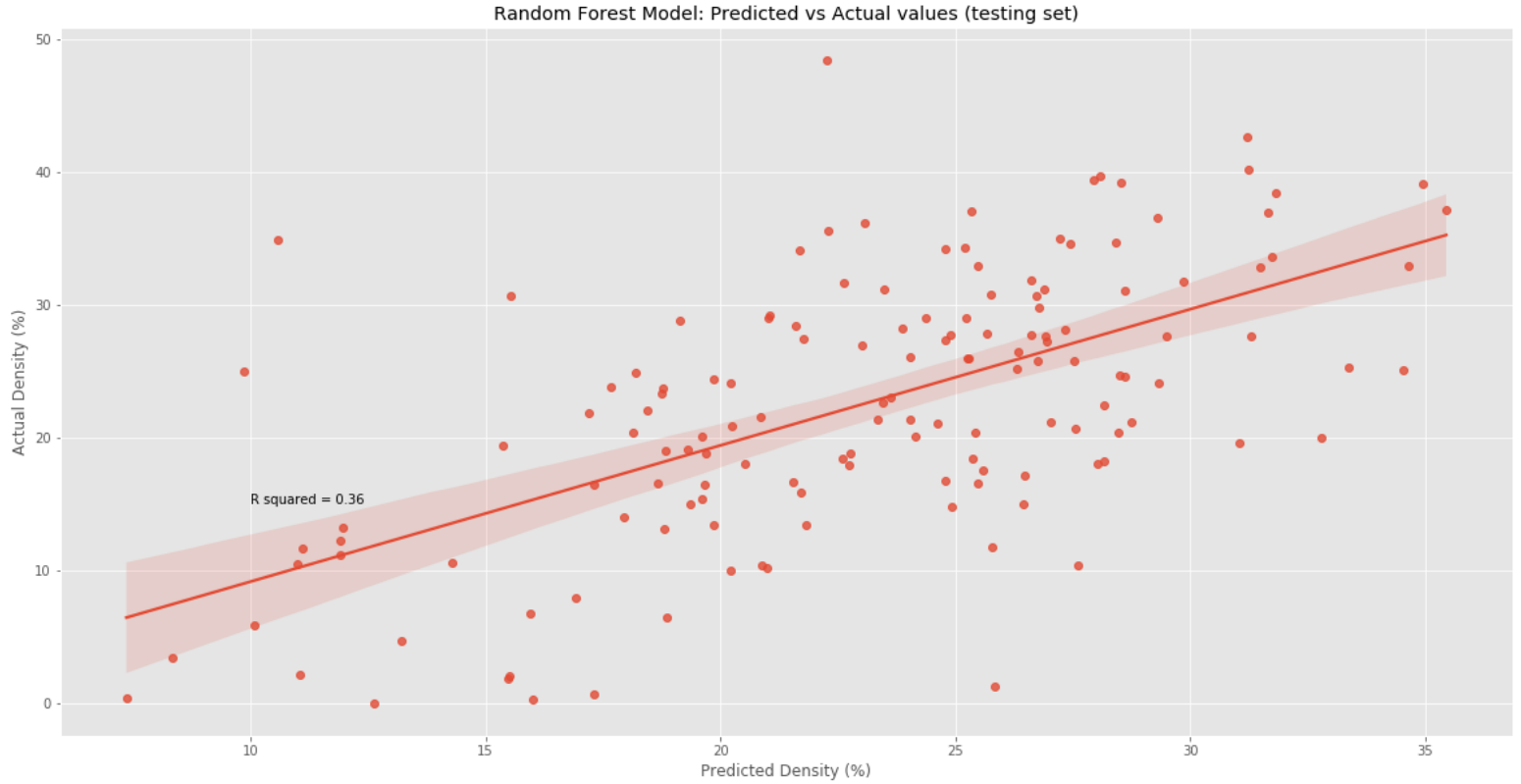


We can indeed predict solar installation rates by LGA using demographics

- Within our final model (using Random Forest Regression), around 36% of the variability in solar installation rates can be predicted by demographic factors

Model	Train set R squared	Test set R squared
Multiple Linear Regression	0.2703166789751422	0.15717867705481747
Boosting Regression (AdaBoost)	0.6810832282921078	0.2939429088986316
Random Forest Regression	0.9222982930546839	0.3645988762013138
Random Forest Regression with Hyperparameter Tuning	1.0	0.3701463941497942

Demographic factors clearly affect solar installation rates



Conclusion and recommendations

- A Random Forest model which was able to explain around 36% of variability in solar installation rates was built
- There is a clear link between solar installation rates in different LGAs and demographics
- Qualitative research should be conducted to elicit what are the most relevant demographic or even individual factors affecting solar installation
- Perform further data collection on a finer scale to create larger datasets for computational analysis, preferably after conducting the qualitative research so data of greater relevance is collected
- Improve upon current algorithm to build a model with higher accuracy

References

- Liu, A. (2020). *Predicting solar installation rates in different locations using demographic data*.
- https://www.pewresearch.org/fact-tank/2016/10/05/americans-strongly-favor-expanding-solar-power-to-help-address-costs-and-environmental-concerns/
- https://www.altestore.com/howto/images/article/world_solar_insolation_data.gif
- Australian PV Institute (APVI) Solar Map, funded by the Australian Renewable Energy Agency, accessed from pv-map.apvi.org.au on 19 February 2020.
- Australian Bureau of Statistics 2020, viewed 26 February 2020, http://stat.data.abs.gov.au.
- https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d